**Problem 1)** Lecture notes contain script network-count.py in both Spark Streaming APIand Spark Structured Streaming API. Use Linux nc (NetCat) utility to demonstrate that scripts work. Run both scripts on your own VM with Spark 2.2 installation. Cloudera VM with Spark 1.6 does not have Spark Structured Streaming API.

**(a) network-count streaming**

[rsha@localhost ~]$ nc -lk 9999

hello

how are you

where have you been

this is cscie63

[rsha@localhost hw6]$ spark-submit --master local[4] network-count.py localhost 9999

-------------------------------------------

Time: 2017-10-14 06:58:27

-------------------------------------------

-------------------------------------------

Time: 2017-10-14 06:58:30

-------------------------------------------

-------------------------------------------

Time: 2017-10-14 06:58:33

-------------------------------------------

-------------------------------------------

Time: 2017-10-14 06:58:36

-------------------------------------------

(u'hello', 1)

-------------------------------------------

Time: 2017-10-14 06:58:39

-------------------------------------------

(u'how', 1)

(u'are', 1)

(u'you', 1)

-------------------------------------------

Time: 2017-10-14 06:58:42

-------------------------------------------

-------------------------------------------

Time: 2017-10-14 06:58:45

-------------------------------------------

-------------------------------------------

Time: 2017-10-14 06:58:48

-------------------------------------------

-------------------------------------------

Time: 2017-10-14 06:58:51

-------------------------------------------

-------------------------------------------

Time: 2017-10-14 06:58:54

-------------------------------------------

(u'where', 1)

(u'you', 1)

(u'have', 1)

(u'been', 1)

-------------------------------------------

Time: 2017-10-14 06:58:57

-------------------------------------------

-------------------------------------------

Time: 2017-10-14 06:59:00

-------------------------------------------

(u'this', 1)

(u'cscie63', 1)

(u'is', 1)

-------------------------------------------

Time: 2017-10-14 06:59:03

-------------------------------------------

-------------------------------------------

Time: 2017-10-14 06:59:06

-------------------------------------------

**(b) network-count structured streaming**

[rsha@localhost input]$ nc -lk 9999

hello

how are you

where have you been

this is cscie63

[rsha@localhost hw6]$ spark-submit --master local[4] ss-network-count.py localhost 9999

-------------------------------------------

Batch: 0

-------------------------------------------

+-----+-----+

| word|count|

+-----+-----+

|hello| 1|

+-----+-----+

-------------------------------------------

Batch: 1

-------------------------------------------

+-------+-----+

| word|count|

+-------+-----+

| you| 2|

|cscie63| 1|

| how| 1|

| hello| 1|

| is| 1|

| have| 1|

| been| 1|

| are| 1|

| where| 1|

| this| 1|

+-------+-----+

**Problem 2)** Expand provide orders.tar.gz file. Also, download shell scrips splitAndSend.original.sh and splitAndSend.sh and the Python script count-buys.py. First run splitAndSend.original.sh and count-buys.py. Record the failure mode of count-buys.py. Simply read the error message produced and tell us what is happening. Then run script splitAndSend.sh and Python program count-buys.py and tell us what the results are. In both cases show use contents of your HDFS directories input, output and staging. The second script splitAndSend.sh is supposed to reduce or eliminate the race condition. You might want to rename HDFS directory output from the first run in order to preserve it’s content. In both cases, show the partial contents of your HDFS directories input, output and staging. In the second run, locate an output file named part-00000 that is not empty and show its content to us. Run these experiments on Cloudera VM. You need HDFS for these programs to run.

**(a) First Run :**

The first run produces the below error message – where it cant find the named file – The script splitAndSend.original does not copy the file to the monitored directory input from another directory in hadoop fs. The generated chunks/files are put straight in the input directory.

*Result :*

Caused by: org.apache.hadoop.ipc.RemoteException(java.io.FileNotFoundException): File does not exist: /user/cloudera/input/chunkbs.\_COPYING\_

at org.apache.hadoop.hdfs.server.namenode.INodeFile.valueOf(INodeFile.java:66)

at org.apache.hadoop.hdfs.server.namenode.INodeFile.valueOf(INodeFile.java:56)

at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.getBlockLocationsInt(FSNamesystem.java:2037)

at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.getBlockLocations(FSNamesystem.java:2007)

at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.getBlockLocations(FSNamesystem.java:1920)

at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.getBlockLocations(NameNodeRpcServer.java:572)

at org.apache.hadoop.hdfs.server.namenode.AuthorizationProviderProxyClientProtocol.getBlockLocations(AuthorizationProviderProxyClientProtocol.java:89)

at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolServerSideTranslatorPB.getBlockLocations(ClientNamenodeProtocolServerSideTranslatorPB.java:365)

at org.apache.hadoop.hdfs.protocol.proto.ClientNamenodeProtocolProtos$ClientNamenodeProtocol$2.callBlockingMethod(ClientNamenodeProtocolProtos.java)

at org.apache.hadoop.ipc.ProtobufRpcEngine$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine.java:617)

at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:1073)

at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2217)

at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2213)

at java.security.AccessController.doPrivileged(Native Method)

at javax.security.auth.Subject.doAs(Subject.java:415)

at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1917)

at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2211)

at org.apache.hadoop.ipc.Client.call(Client.java:1502)

at org.apache.hadoop.ipc.Client.call(Client.java:1439)

at org.apache.hadoop.ipc.ProtobufRpcEngine$Invoker.invoke(ProtobufRpcEngine.java:230)

at com.sun.proxy.$Proxy26.getBlockLocations(Unknown Source)

at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolTranslatorPB.getBlockLocations(ClientNamenodeProtocolTranslatorPB.java:256)

at sun.reflect.GeneratedMethodAccessor46.invoke(Unknown Source)

at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)

at java.lang.reflect.Method.invoke(Method.java:606)

at org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvocationHandler.java:260)

at org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocationHandler.java:104)

at com.sun.proxy.$Proxy27.getBlockLocations(Unknown Source)

at org.apache.hadoop.hdfs.DFSClient.callGetBlockLocations(DFSClient.java:1282)

... 36 more

at org.apache.spark.streaming.api.python.TransformFunction.callPythonTransformFunction(PythonDStream.scala:95)

at org.apache.spark.streaming.api.python.TransformFunction.apply(PythonDStream.scala:78)

at org.apache.spark.streaming.api.python.PythonDStream$$anonfun$callForeachRDD$1.apply(PythonDStream.scala:189)

at org.apache.spark.streaming.api.python.PythonDStream$$anonfun$callForeachRDD$1.apply(PythonDStream.scala:189)

at org.apache.spark.streaming.dstream.ForEachDStream$$anonfun$1$$anonfun$apply$mcV$sp$1.apply$mcV$sp(ForEachDStream.scala:50)

at org.apache.spark.streaming.dstream.ForEachDStream$$anonfun$1$$anonfun$apply$mcV$sp$1.apply(ForEachDStream.scala:50)

at org.apache.spark.streaming.dstream.ForEachDStream$$anonfun$1$$anonfun$apply$mcV$sp$1.apply(ForEachDStream.scala:50)

at org.apache.spark.streaming.dstream.DStream.createRDDWithLocalProperties(DStream.scala:426)

at org.apache.spark.streaming.dstream.ForEachDStream$$anonfun$1.apply$mcV$sp(ForEachDStream.scala:49)

at org.apache.spark.streaming.dstream.ForEachDStream$$anonfun$1.apply(ForEachDStream.scala:49)

at org.apache.spark.streaming.dstream.ForEachDStream$$anonfun$1.apply(ForEachDStream.scala:49)

at scala.util.Try$.apply(Try.scala:161)

at org.apache.spark.streaming.scheduler.Job.run(Job.scala:39)

at org.apache.spark.streaming.scheduler.JobScheduler$JobHandler$$anonfun$run$1.apply$mcV$sp(JobScheduler.scala:224)

at org.apache.spark.streaming.scheduler.JobScheduler$JobHandler$$anonfun$run$1.apply(JobScheduler.scala:224)

at org.apache.spark.streaming.scheduler.JobScheduler$JobHandler$$anonfun$run$1.apply(JobScheduler.scala:224)

at scala.util.DynamicVariable.withValue(DynamicVariable.scala:57)

at org.apache.spark.streaming.scheduler.JobScheduler$JobHandler.run(JobScheduler.scala:223)

at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1145)

at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:615)

at java.lang.Thread.run(Thread.java:745)

*Contents of input , output and staging in first run:*

[cloudera@quickstart assignment6]$ hadoop fs -ls -R input

-rw-r--r-- 1 cloudera cloudera 437626 2017-10-10 19:27 input/chunkaa

-rw-r--r-- 1 cloudera cloudera 448647 2017-10-10 19:27 input/chunkab

-rw-r--r-- 1 cloudera cloudera 448605 2017-10-10 19:27 input/chunkac

-rw-r--r-- 1 cloudera cloudera 448794 2017-10-10 19:27 input/chunkad

-rw-r--r-- 1 cloudera cloudera 448624 2017-10-10 19:27 input/chunkae

-rw-r--r-- 1 cloudera cloudera 448553 2017-10-10 19:28 input/chunkaf

-rw-r--r-- 1 cloudera cloudera 448436 2017-10-10 19:28 input/chunkag

-rw-r--r-- 1 cloudera cloudera 448679 2017-10-10 19:28 input/chunkah

-rw-r--r-- 1 cloudera cloudera 448424 2017-10-10 19:28 input/chunkai

-rw-r--r-- 1 cloudera cloudera 448564 2017-10-10 19:28 input/chunkaj

-rw-r--r-- 1 cloudera cloudera 458595 2017-10-10 19:28 input/chunkak

-rw-r--r-- 1 cloudera cloudera 458580 2017-10-10 19:28 input/chunkal

-rw-r--r-- 1 cloudera cloudera 458605 2017-10-10 19:28 input/chunkam

-rw-r--r-- 1 cloudera cloudera 458630 2017-10-10 19:28 input/chunkan

-rw-r--r-- 1 cloudera cloudera 458663 2017-10-10 19:29 input/chunkao

-rw-r--r-- 1 cloudera cloudera 458540 2017-10-10 19:29 input/chunkap

-rw-r--r-- 1 cloudera cloudera 458533 2017-10-10 19:29 input/chunkaq

-rw-r--r-- 1 cloudera cloudera 458403 2017-10-10 19:29 input/chunkar

-rw-r--r-- 1 cloudera cloudera 458470 2017-10-10 19:29 input/chunkas

-rw-r--r-- 1 cloudera cloudera 458584 2017-10-10 19:29 input/chunkat

-rw-r--r-- 1 cloudera cloudera 458333 2017-10-10 19:29 input/chunkau

-rw-r--r-- 1 cloudera cloudera 458600 2017-10-10 19:29 input/chunkav

-rw-r--r-- 1 cloudera cloudera 458732 2017-10-10 19:30 input/chunkaw

-rw-r--r-- 1 cloudera cloudera 458559 2017-10-10 19:30 input/chunkax

-rw-r--r-- 1 cloudera cloudera 458654 2017-10-10 19:30 input/chunkay

-rw-r--r-- 1 cloudera cloudera 458533 2017-10-10 19:30 input/chunkaz

-rw-r--r-- 1 cloudera cloudera 458535 2017-10-10 19:30 input/chunkba

-rw-r--r-- 1 cloudera cloudera 458700 2017-10-10 19:30 input/chunkbb

-rw-r--r-- 1 cloudera cloudera 458509 2017-10-10 19:30 input/chunkbc

-rw-r--r-- 1 cloudera cloudera 458488 2017-10-10 19:30 input/chunkbd

-rw-r--r-- 1 cloudera cloudera 458527 2017-10-10 19:30 input/chunkbe

-rw-r--r-- 1 cloudera cloudera 458662 2017-10-10 19:30 input/chunkbf

-rw-r--r-- 1 cloudera cloudera 458618 2017-10-10 19:31 input/chunkbg

-rw-r--r-- 1 cloudera cloudera 458673 2017-10-10 19:31 input/chunkbh

-rw-r--r-- 1 cloudera cloudera 458431 2017-10-10 19:31 input/chunkbi

-rw-r--r-- 1 cloudera cloudera 458464 2017-10-10 19:31 input/chunkbj

-rw-r--r-- 1 cloudera cloudera 458450 2017-10-10 19:31 input/chunkbk

-rw-r--r-- 1 cloudera cloudera 458536 2017-10-10 19:31 input/chunkbl

-rw-r--r-- 1 cloudera cloudera 458412 2017-10-10 19:31 input/chunkbm

-rw-r--r-- 1 cloudera cloudera 458712 2017-10-10 19:31 input/chunkbn

-rw-r--r-- 1 cloudera cloudera 458586 2017-10-10 19:31 input/chunkbo

-rw-r--r-- 1 cloudera cloudera 458501 2017-10-10 19:32 input/chunkbp

-rw-r--r-- 1 cloudera cloudera 458569 2017-10-10 19:32 input/chunkbq

-rw-r--r-- 1 cloudera cloudera 458459 2017-10-10 19:32 input/chunkbr

-rw-r--r-- 1 cloudera cloudera 458737 2017-10-10 19:32 input/chunkbs

-rw-r--r-- 1 cloudera cloudera 458556 2017-10-10 19:32 input/chunkbt

-rw-r--r-- 1 cloudera cloudera 458358 2017-10-10 19:32 input/chunkbu

-rw-r--r-- 1 cloudera cloudera 458570 2017-10-10 19:32 input/chunkbv

-rw-r--r-- 1 cloudera cloudera 458639 2017-10-10 19:32 input/chunkbw

-rw-r--r-- 1 cloudera cloudera 458503 2017-10-10 19:32 input/chunkbx

[cloudera@quickstart assignment6]$ hadoop fs -ls -R output

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:26 output/output-1507688811000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:26 output/output-1507688811000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:26 output/output-1507688811000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:27 output/output-1507688820000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:27 output/output-1507688820000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:27 output/output-1507688820000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:27 output/output-1507688829000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:27 output/output-1507688829000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:27 output/output-1507688829000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:27 output/output-1507688838000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:27 output/output-1507688838000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:27 output/output-1507688838000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:27 output/output-1507688847000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:27 output/output-1507688847000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:27 output/output-1507688847000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:27 output/output-1507688856000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:27 output/output-1507688856000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:27 output/output-1507688856000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:27 output/output-1507688865000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:27 output/output-1507688865000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 30 2017-10-10 19:27 output/output-1507688865000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:27 output/output-1507688874000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:27 output/output-1507688874000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:27 output/output-1507688874000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:28 output/output-1507688883000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:28 output/output-1507688883000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:28 output/output-1507688883000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:28 output/output-1507688892000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:28 output/output-1507688892000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:28 output/output-1507688892000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:28 output/output-1507688901000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:28 output/output-1507688901000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:28 output/output-1507688901000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:28 output/output-1507688910000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:28 output/output-1507688910000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:28 output/output-1507688910000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:28 output/output-1507688919000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:28 output/output-1507688919000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 30 2017-10-10 19:28 output/output-1507688919000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:28 output/output-1507688928000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:28 output/output-1507688928000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:28 output/output-1507688928000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:28 output/output-1507688937000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:28 output/output-1507688937000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:28 output/output-1507688937000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:29 output/output-1507688946000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:29 output/output-1507688946000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:29 output/output-1507688946000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:29 output/output-1507688955000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:29 output/output-1507688955000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 30 2017-10-10 19:29 output/output-1507688955000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:29 output/output-1507688964000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:29 output/output-1507688964000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:29 output/output-1507688964000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:29 output/output-1507688973000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:29 output/output-1507688973000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:29 output/output-1507688973000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:29 output/output-1507688982000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:29 output/output-1507688982000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 30 2017-10-10 19:29 output/output-1507688982000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:29 output/output-1507688991000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:29 output/output-1507688991000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:29 output/output-1507688991000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:30 output/output-1507689000000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:30 output/output-1507689000000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:30 output/output-1507689000000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:30 output/output-1507689009000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:30 output/output-1507689009000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 30 2017-10-10 19:30 output/output-1507689009000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:30 output/output-1507689018000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:30 output/output-1507689018000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:30 output/output-1507689018000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:30 output/output-1507689027000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:30 output/output-1507689027000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 30 2017-10-10 19:30 output/output-1507689027000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:30 output/output-1507689036000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:30 output/output-1507689036000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:30 output/output-1507689036000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:30 output/output-1507689045000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:30 output/output-1507689045000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:30 output/output-1507689045000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:30 output/output-1507689054000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:30 output/output-1507689054000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 30 2017-10-10 19:30 output/output-1507689054000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:31 output/output-1507689063000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:31 output/output-1507689063000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:31 output/output-1507689063000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:31 output/output-1507689072000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:31 output/output-1507689072000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:31 output/output-1507689072000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:31 output/output-1507689081000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:31 output/output-1507689081000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 30 2017-10-10 19:31 output/output-1507689081000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:31 output/output-1507689090000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:31 output/output-1507689090000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:31 output/output-1507689090000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:31 output/output-1507689099000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:31 output/output-1507689099000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 30 2017-10-10 19:31 output/output-1507689099000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:31 output/output-1507689108000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:31 output/output-1507689108000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:31 output/output-1507689108000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:31 output/output-1507689117000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:31 output/output-1507689117000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:31 output/output-1507689117000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:32 output/output-1507689126000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:32 output/output-1507689126000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 30 2017-10-10 19:32 output/output-1507689126000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:32 output/output-1507689135000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:32 output/output-1507689135000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:32 output/output-1507689135000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:32 output/output-1507689144000.txt

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:32 output/output-1507689144000.txt/\_temporary

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:32 output/output-1507689144000.txt/\_temporary/0

[cloudera@quickstart assignment6]$ hadoop fs -ls -R staging

[cloudera@quickstart assignment6]$

**(b) Second run**

*setup of input, output directories to ensure they are clean :*

183 hadoop fs -mv output output\_run1

184 hadoop fs -ls

185 hadoop fs -mkdir output

186 hadoop fs -ls input

[cloudera@quickstart assignment6]$ hadoop fs -mv input input\_run1

[cloudera@quickstart assignment6]$ hadoop fs -mkdir input

*Contents of input, output and staging directories after second run :*

[cloudera@quickstart assignment6]$ hadoop fs -ls input

Found 50 items

-rw-r--r-- 1 cloudera cloudera 437626 2017-10-10 19:52 input/chunkaa

-rw-r--r-- 1 cloudera cloudera 448647 2017-10-10 19:52 input/chunkab

-rw-r--r-- 1 cloudera cloudera 448605 2017-10-10 19:52 input/chunkac

-rw-r--r-- 1 cloudera cloudera 448794 2017-10-10 19:52 input/chunkad

-rw-r--r-- 1 cloudera cloudera 448624 2017-10-10 19:53 input/chunkae

-rw-r--r-- 1 cloudera cloudera 448553 2017-10-10 19:53 input/chunkaf

-rw-r--r-- 1 cloudera cloudera 448436 2017-10-10 19:53 input/chunkag

-rw-r--r-- 1 cloudera cloudera 448679 2017-10-10 19:53 input/chunkah

-rw-r--r-- 1 cloudera cloudera 448424 2017-10-10 19:53 input/chunkai

-rw-r--r-- 1 cloudera cloudera 448564 2017-10-10 19:53 input/chunkaj

-rw-r--r-- 1 cloudera cloudera 458595 2017-10-10 19:54 input/chunkak

-rw-r--r-- 1 cloudera cloudera 458580 2017-10-10 19:54 input/chunkal

-rw-r--r-- 1 cloudera cloudera 458605 2017-10-10 19:54 input/chunkam

-rw-r--r-- 1 cloudera cloudera 458630 2017-10-10 19:54 input/chunkan

-rw-r--r-- 1 cloudera cloudera 458663 2017-10-10 19:54 input/chunkao

-rw-r--r-- 1 cloudera cloudera 458540 2017-10-10 19:54 input/chunkap

-rw-r--r-- 1 cloudera cloudera 458533 2017-10-10 19:55 input/chunkaq

-rw-r--r-- 1 cloudera cloudera 458403 2017-10-10 19:55 input/chunkar

-rw-r--r-- 1 cloudera cloudera 458470 2017-10-10 19:55 input/chunkas

-rw-r--r-- 1 cloudera cloudera 458584 2017-10-10 19:55 input/chunkat

-rw-r--r-- 1 cloudera cloudera 458333 2017-10-10 19:55 input/chunkau

-rw-r--r-- 1 cloudera cloudera 458600 2017-10-10 19:55 input/chunkav

-rw-r--r-- 1 cloudera cloudera 458732 2017-10-10 19:56 input/chunkaw

-rw-r--r-- 1 cloudera cloudera 458559 2017-10-10 19:56 input/chunkax

-rw-r--r-- 1 cloudera cloudera 458654 2017-10-10 19:56 input/chunkay

-rw-r--r-- 1 cloudera cloudera 458533 2017-10-10 19:56 input/chunkaz

-rw-r--r-- 1 cloudera cloudera 458535 2017-10-10 19:56 input/chunkba

-rw-r--r-- 1 cloudera cloudera 458700 2017-10-10 19:56 input/chunkbb

-rw-r--r-- 1 cloudera cloudera 458509 2017-10-10 19:57 input/chunkbc

-rw-r--r-- 1 cloudera cloudera 458488 2017-10-10 19:57 input/chunkbd

-rw-r--r-- 1 cloudera cloudera 458527 2017-10-10 19:57 input/chunkbe

-rw-r--r-- 1 cloudera cloudera 458662 2017-10-10 19:57 input/chunkbf

-rw-r--r-- 1 cloudera cloudera 458618 2017-10-10 19:57 input/chunkbg

-rw-r--r-- 1 cloudera cloudera 458673 2017-10-10 19:57 input/chunkbh

-rw-r--r-- 1 cloudera cloudera 458431 2017-10-10 19:58 input/chunkbi

-rw-r--r-- 1 cloudera cloudera 458464 2017-10-10 19:58 input/chunkbj

-rw-r--r-- 1 cloudera cloudera 458450 2017-10-10 19:58 input/chunkbk

-rw-r--r-- 1 cloudera cloudera 458536 2017-10-10 19:58 input/chunkbl

-rw-r--r-- 1 cloudera cloudera 458412 2017-10-10 19:58 input/chunkbm

-rw-r--r-- 1 cloudera cloudera 458712 2017-10-10 19:58 input/chunkbn

-rw-r--r-- 1 cloudera cloudera 458586 2017-10-10 19:59 input/chunkbo

-rw-r--r-- 1 cloudera cloudera 458501 2017-10-10 19:59 input/chunkbp

-rw-r--r-- 1 cloudera cloudera 458569 2017-10-10 19:59 input/chunkbq

-rw-r--r-- 1 cloudera cloudera 458459 2017-10-10 19:59 input/chunkbr

-rw-r--r-- 1 cloudera cloudera 458737 2017-10-10 19:59 input/chunkbs

-rw-r--r-- 1 cloudera cloudera 458556 2017-10-10 19:59 input/chunkbt

-rw-r--r-- 1 cloudera cloudera 458358 2017-10-10 20:00 input/chunkbu

-rw-r--r-- 1 cloudera cloudera 458570 2017-10-10 20:00 input/chunkbv

-rw-r--r-- 1 cloudera cloudera 458639 2017-10-10 20:00 input/chunkbw

-rw-r--r-- 1 cloudera cloudera 458503 2017-10-10 20:00 input/chunkbx

[cloudera@quickstart assignment6]$ hadoop fs -ls staging

[cloudera@quickstart assignment6]$

[cloudera@quickstart assignment6]$ hadoop fs -ls -R output

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:52 output/output-1507690341000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:52 output/output-1507690341000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:52 output/output-1507690341000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:52 output/output-1507690350000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:52 output/output-1507690350000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:52 output/output-1507690350000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:52 output/output-1507690359000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:52 output/output-1507690359000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:52 output/output-1507690359000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:52 output/output-1507690368000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:52 output/output-1507690368000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:52 output/output-1507690368000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:52 output/output-1507690377000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:52 output/output-1507690377000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:52 output/output-1507690377000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:53 output/output-1507690386000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:53 output/output-1507690386000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:53 output/output-1507690386000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:53 output/output-1507690395000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:53 output/output-1507690395000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:53 output/output-1507690395000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:53 output/output-1507690404000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:53 output/output-1507690404000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:53 output/output-1507690404000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:53 output/output-1507690413000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:53 output/output-1507690413000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:53 output/output-1507690413000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:53 output/output-1507690422000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:53 output/output-1507690422000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:53 output/output-1507690422000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:53 output/output-1507690431000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:53 output/output-1507690431000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:53 output/output-1507690431000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:54 output/output-1507690440000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:54 output/output-1507690440000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:54 output/output-1507690440000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:54 output/output-1507690449000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:54 output/output-1507690449000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:54 output/output-1507690449000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:54 output/output-1507690458000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:54 output/output-1507690458000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:54 output/output-1507690458000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:54 output/output-1507690467000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:54 output/output-1507690467000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:54 output/output-1507690467000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:54 output/output-1507690476000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:54 output/output-1507690476000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:54 output/output-1507690476000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:54 output/output-1507690485000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:54 output/output-1507690485000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:54 output/output-1507690485000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:54 output/output-1507690494000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:54 output/output-1507690494000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:54 output/output-1507690494000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:55 output/output-1507690503000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:55 output/output-1507690503000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:55 output/output-1507690503000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:55 output/output-1507690512000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:55 output/output-1507690512000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:55 output/output-1507690512000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:55 output/output-1507690521000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:55 output/output-1507690521000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:55 output/output-1507690521000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:55 output/output-1507690530000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:55 output/output-1507690530000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:55 output/output-1507690530000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:55 output/output-1507690539000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:55 output/output-1507690539000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:55 output/output-1507690539000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:55 output/output-1507690548000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:55 output/output-1507690548000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:55 output/output-1507690548000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:55 output/output-1507690557000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:55 output/output-1507690557000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:55 output/output-1507690557000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:56 output/output-1507690566000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:56 output/output-1507690566000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:56 output/output-1507690566000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:56 output/output-1507690575000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:56 output/output-1507690575000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:56 output/output-1507690575000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:56 output/output-1507690584000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:56 output/output-1507690584000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:56 output/output-1507690584000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:56 output/output-1507690593000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:56 output/output-1507690593000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:56 output/output-1507690593000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:56 output/output-1507690602000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:56 output/output-1507690602000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:56 output/output-1507690602000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:56 output/output-1507690611000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:56 output/output-1507690611000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:56 output/output-1507690611000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:57 output/output-1507690620000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:57 output/output-1507690620000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:57 output/output-1507690620000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:57 output/output-1507690629000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:57 output/output-1507690629000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:57 output/output-1507690629000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:57 output/output-1507690638000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:57 output/output-1507690638000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:57 output/output-1507690638000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:57 output/output-1507690647000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:57 output/output-1507690647000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:57 output/output-1507690647000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:57 output/output-1507690656000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:57 output/output-1507690656000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:57 output/output-1507690656000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:57 output/output-1507690665000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:57 output/output-1507690665000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:57 output/output-1507690665000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:57 output/output-1507690674000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:57 output/output-1507690674000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:57 output/output-1507690674000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:58 output/output-1507690683000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:58 output/output-1507690683000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:58 output/output-1507690683000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:58 output/output-1507690692000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:58 output/output-1507690692000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:58 output/output-1507690692000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:58 output/output-1507690701000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:58 output/output-1507690701000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:58 output/output-1507690701000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:58 output/output-1507690710000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:58 output/output-1507690710000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:58 output/output-1507690710000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:58 output/output-1507690719000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:58 output/output-1507690719000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:58 output/output-1507690719000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:58 output/output-1507690728000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:58 output/output-1507690728000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:58 output/output-1507690728000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:58 output/output-1507690737000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:58 output/output-1507690737000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:58 output/output-1507690737000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:59 output/output-1507690746000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:59 output/output-1507690746000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:59 output/output-1507690746000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:59 output/output-1507690755000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:59 output/output-1507690755000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:59 output/output-1507690755000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:59 output/output-1507690764000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:59 output/output-1507690764000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:59 output/output-1507690764000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:59 output/output-1507690773000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:59 output/output-1507690773000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:59 output/output-1507690773000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:59 output/output-1507690782000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:59 output/output-1507690782000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:59 output/output-1507690782000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 19:59 output/output-1507690791000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 19:59 output/output-1507690791000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 19:59 output/output-1507690791000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 20:00 output/output-1507690800000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:00 output/output-1507690800000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 20:00 output/output-1507690800000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 20:00 output/output-1507690809000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:00 output/output-1507690809000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 20:00 output/output-1507690809000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 20:00 output/output-1507690818000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:00 output/output-1507690818000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 20:00 output/output-1507690818000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 20:00 output/output-1507690827000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:00 output/output-1507690827000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 20:00 output/output-1507690827000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 20:00 output/output-1507690836000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:00 output/output-1507690836000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:00 output/output-1507690836000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 20:00 output/output-1507690845000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:00 output/output-1507690845000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 20:00 output/output-1507690845000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 20:00 output/output-1507690854000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:00 output/output-1507690854000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 29 2017-10-10 20:00 output/output-1507690854000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 20:01 output/output-1507690863000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:01 output/output-1507690863000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:01 output/output-1507690863000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 20:01 output/output-1507690872000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:01 output/output-1507690872000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:01 output/output-1507690872000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 20:01 output/output-1507690881000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:01 output/output-1507690881000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:01 output/output-1507690881000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 20:01 output/output-1507690890000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:01 output/output-1507690890000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:01 output/output-1507690890000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 20:01 output/output-1507690899000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:01 output/output-1507690899000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:01 output/output-1507690899000.txt/part-00000

drwxr-xr-x - cloudera cloudera 0 2017-10-10 20:01 output/output-1507690908000.txt

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:01 output/output-1507690908000.txt/\_SUCCESS

-rw-r--r-- 1 cloudera cloudera 0 2017-10-10 20:01 output/output-1507690908000.txt/part-00000

*Contents of a couple of non empty part-0000 file*

[cloudera@quickstart assignment6]$ hadoop fs -cat output/output-1507690350000.txt/part-00000

(False, 4980L)

(True, 5020L)

[cloudera@quickstart assignment6]$ hadoop fs -cat output/output-1507690854000.txt/part-00000

(False, 4989L)

(True, 5011L)

**Problem 3).** In the second run of the previous problem you will notice that many of part-00000 files in your output directory are empty. Could you explain why.

(10%)

Count-buys.py is set up with a batchduration of 9 sec – which means that it expects a stream to arrive every 9 sec. The splitAndSend sctipt sleeps for 3 sec, after which it moves the file from staging to input directory. Due to this setup it is possible that the count buys program (count-buys.py) looks in the input directory finds no file(s) , and because the program ( count-buys.py) writes to an output file in each run, it produces an empty output file in case where there are no files in the input directory.

**Problem 4)** Could you rewrite count-buys.sh in Spark Structured Streaming API. If you do that change script splitAndSend.sh to move generated chunks from the local files system directory staging to local file system directory input. Run this experiment on your VM with Spark 2.2.

(20%)

1. *modified splitAndSend.local.sh*

#!/bin/bash

if [ -z "$1" ]; then

echo "Missing output folder name"

exit 1

fi

split -l 10000 orders.txt chunk

for f in `ls chunk\*`; do

mv $f /home/rsha/staging

sleep 3

mv /home/rsha/staging/$f /home/rsha/$1/

rm -f $f

done

1. *Structure Streaming count-buys-ss.py*

from pyspark.sql import SparkSession

from pyspark.sql.types import StructType

spark = SparkSession.builder.appName("SparkStructuredStreamingCountBuys").getOrCreate()

structType = StructType().add("time", "string").add("oId", "integer").add("cId", "integer")\

.add("syn", "string") \

.add("qty", "integer") \

.add("price", "float") \

.add("buy", "string")

filestream = spark.readStream.csv("/home/rsha/input", schema=structType)

buySellCount = filestream.groupBy("buy").count()

query = buySellCount.writeStream.format("console").outputMode("complete") \

.start()

query.awaitTermination()

# ssc.stop(False)

(c ) *Run and results from run:*

rsha@localhost hw6]$ spark-submit --master local[4] count-buys-ss.py localhost 9999

-------------------------------------------

Batch: 0

-------------------------------------------

+---+-----+

|buy|count|

+---+-----+

| B| 5020|

| S| 4980|

+---+-----+

-------------------------------------------

Batch: 1

-------------------------------------------

+---+-----+

|buy|count|

+---+-----+

| B|25182|

| S|24818|

+---+-----+

-------------------------------------------

Batch: 2

-------------------------------------------

+---+-----+

|buy|count|

+---+-----+

| B|35213|

| S|34787|

+---+-----+

-------------------------------------------

Batch: 3

-------------------------------------------

+---+-----+

|buy|count|

+---+-----+

| B|45282|

| S|44718|

+---+-----+

-------------------------------------------

Batch: 4

-------------------------------------------

+---+-----+

|buy|count|

+---+-----+

| B|50313|

| S|49687|

+---+-----+

-------------------------------------------

Batch: 5

-------------------------------------------

+---+-----+

|buy|count|

+---+-----+

| B|60431|

| S|59569|

+---+-----+

-------------------------------------------

Batch: 6

-------------------------------------------

+---+-----+

|buy|count|

+---+-----+

| B|70483|

| S|69517|

+---+-----+

-------------------------------------------

Batch: 7

-------------------------------------------

+---+-----+

|buy|count|

+---+-----+

| B|80508|

| S|79492|

+---+-----+

*input directory*

[rsha@localhost input]$ ls -ltr

total 22308

-rw-rw-r--. 1 rsha rsha 437626 Oct 12 20:19 chunkaa

-rw-rw-r--. 1 rsha rsha 448647 Oct 12 20:19 chunkab

-rw-rw-r--. 1 rsha rsha 448605 Oct 12 20:19 chunkac

-rw-rw-r--. 1 rsha rsha 448624 Oct 12 20:19 chunkae

-rw-rw-r--. 1 rsha rsha 448794 Oct 12 20:19 chunkad

-rw-rw-r--. 1 rsha rsha 448553 Oct 12 20:19 chunkaf

-rw-rw-r--. 1 rsha rsha 448436 Oct 12 20:19 chunkag

-rw-rw-r--. 1 rsha rsha 448424 Oct 12 20:19 chunkai

-rw-rw-r--. 1 rsha rsha 448679 Oct 12 20:19 chunkah

-rw-rw-r--. 1 rsha rsha 458595 Oct 12 20:19 chunkak

-rw-rw-r--. 1 rsha rsha 448564 Oct 12 20:19 chunkaj

-rw-rw-r--. 1 rsha rsha 458580 Oct 12 20:19 chunkal

-rw-rw-r--. 1 rsha rsha 458605 Oct 12 20:19 chunkam

-rw-rw-r--. 1 rsha rsha 458663 Oct 12 20:19 chunkao

-rw-rw-r--. 1 rsha rsha 458630 Oct 12 20:19 chunkan

-rw-rw-r--. 1 rsha rsha 458533 Oct 12 20:19 chunkaq

-rw-rw-r--. 1 rsha rsha 458540 Oct 12 20:19 chunkap

-rw-rw-r--. 1 rsha rsha 458403 Oct 12 20:19 chunkar

-rw-rw-r--. 1 rsha rsha 458470 Oct 12 20:19 chunkas

-rw-rw-r--. 1 rsha rsha 458333 Oct 12 20:19 chunkau

-rw-rw-r--. 1 rsha rsha 458584 Oct 12 20:19 chunkat

-rw-rw-r--. 1 rsha rsha 458600 Oct 12 20:19 chunkav

-rw-rw-r--. 1 rsha rsha 458559 Oct 12 20:19 chunkax

-rw-rw-r--. 1 rsha rsha 458732 Oct 12 20:19 chunkaw

-rw-rw-r--. 1 rsha rsha 458533 Oct 12 20:19 chunkaz

-rw-rw-r--. 1 rsha rsha 458654 Oct 12 20:19 chunkay

-rw-rw-r--. 1 rsha rsha 458535 Oct 12 20:19 chunkba

-rw-rw-r--. 1 rsha rsha 458509 Oct 12 20:19 chunkbc

-rw-rw-r--. 1 rsha rsha 458700 Oct 12 20:19 chunkbb

-rw-rw-r--. 1 rsha rsha 458488 Oct 12 20:19 chunkbd

-rw-rw-r--. 1 rsha rsha 458527 Oct 12 20:19 chunkbe

-rw-rw-r--. 1 rsha rsha 458618 Oct 12 20:19 chunkbg

-rw-rw-r--. 1 rsha rsha 458662 Oct 12 20:19 chunkbf

-rw-rw-r--. 1 rsha rsha 458673 Oct 12 20:19 chunkbh

-rw-rw-r--. 1 rsha rsha 458431 Oct 12 20:19 chunkbi

-rw-rw-r--. 1 rsha rsha 458450 Oct 12 20:19 chunkbk

-rw-rw-r--. 1 rsha rsha 458464 Oct 12 20:19 chunkbj

-rw-rw-r--. 1 rsha rsha 458412 Oct 12 20:19 chunkbm

-rw-rw-r--. 1 rsha rsha 458536 Oct 12 20:19 chunkbl

-rw-rw-r--. 1 rsha rsha 458712 Oct 12 20:19 chunkbn

-rw-rw-r--. 1 rsha rsha 458501 Oct 12 20:19 chunkbp

-rw-rw-r--. 1 rsha rsha 458586 Oct 12 20:19 chunkbo

-rw-rw-r--. 1 rsha rsha 458569 Oct 12 20:19 chunkbq

-rw-rw-r--. 1 rsha rsha 458737 Oct 12 20:19 chunkbs

-rw-rw-r--. 1 rsha rsha 458459 Oct 12 20:19 chunkbr

-rw-rw-r--. 1 rsha rsha 458358 Oct 12 20:19 chunkbu

-rw-rw-r--. 1 rsha rsha 458556 Oct 12 20:19 chunkbt

-rw-rw-r--. 1 rsha rsha 458570 Oct 12 20:19 chunkbv

-rw-rw-r--. 1 rsha rsha 458503 Oct 12 20:19 chunkbx

-rw-rw-r--. 1 rsha rsha 458639 Oct 12 20:19 chunkbw

**Problem 5)** Examine provided Python program stateful\_wordcount.py. Make it work as is. If there are errors on the code, fix them. Modify the code so that it outputs the number of words starting with letters a and b. Demonstrate that modified program work. You should provide several both positive and negative examples.

(20%)

1. *stateful\_wordcount run*

*Result :*

[rsha@localhost input]$ nc -lk 9999

hello

how are you

good to see you again

lets go get some coffee

[rsha@localhost hw6]$ spark-submit --master local[4] stateful\_wordcount.py localhost 9999

-------------------------------------------

Time: 2017-10-12 21:08:24

-------------------------------------------

-------------------------------------------

Time: 2017-10-12 21:08:25

-------------------------------------------

-------------------------------------------

Time: 2017-10-12 21:08:26

-------------------------------------------

(u'hello', 1)

-------------------------------------------

Time: 2017-10-12 21:08:27

-------------------------------------------

(u'how', 1)

(u'hello', 1)

(u'are', 1)

(u'you', 1)

-------------------------------------------

Time: 2017-10-12 21:08:28

-------------------------------------------

(u'how', 1)

(u'hello', 1)

(u'are', 1)

(u'you', 1)

-------------------------------------------

Time: 2017-10-12 21:08:29

-------------------------------------------

(u'how', 1)

(u'hello', 1)

(u'are', 1)

(u'you', 1)

-------------------------------------------

Time: 2017-10-12 21:08:30

-------------------------------------------

(u'how', 1)

(u'hello', 1)

(u'are', 1)

(u'you', 1)

-------------------------------------------

Time: 2017-10-12 21:08:31

-------------------------------------------

(u'how', 1)

(u'hello', 1)

(u'are', 1)

(u'you', 1)

-------------------------------------------

Time: 2017-10-12 21:08:32

-------------------------------------------

(u'are', 1)

(u'how', 1)

(u'hello', 1)

(u'to', 1)

(u'again', 1)

(u'you', 2)

(u'see', 1)

(u'good', 1)

-------------------------------------------

Time: 2017-10-12 21:08:33

-------------------------------------------

(u'how', 1)

(u'hello', 1)

(u'are', 1)

(u'to', 1)

(u'again', 1)

(u'you', 2)

(u'see', 1)

(u'good', 1)

-------------------------------------------

Time: 2017-10-12 21:08:34

-------------------------------------------

(u'are', 1)

(u'how', 1)

(u'hello', 1)

(u'to', 1)

(u'again', 1)

(u'you', 2)

(u'see', 1)

(u'good', 1)

-------------------------------------------

Time: 2017-10-12 21:08:35

-------------------------------------------

(u'how', 1)

(u'hello', 1)

(u'are', 1)

(u'to', 1)

(u'again', 1)

(u'you', 2)

(u'see', 1)

(u'good', 1)

-------------------------------------------

Time: 2017-10-12 21:08:36

-------------------------------------------

(u'are', 1)

(u'how', 1)

(u'hello', 1)

(u'to', 1)

(u'again', 1)

(u'you', 2)

(u'see', 1)

(u'good', 1)

-------------------------------------------

Time: 2017-10-12 21:08:37

-------------------------------------------

(u'how', 1)

(u'hello', 1)

(u'are', 1)

(u'to', 1)

(u'again', 1)

(u'you', 2)

(u'see', 1)

(u'good', 1)

-------------------------------------------

Time: 2017-10-12 21:08:38

-------------------------------------------

(u'go', 1)

(u'some', 1)

(u'are', 1)

(u'how', 1)

(u'hello', 1)

(u'to', 1)

(u'again', 1)

(u'you', 2)

(u'coffee', 1)

(u'lets', 1)

...

-------------------------------------------

Time: 2017-10-12 21:08:39

-------------------------------------------

(u'go', 1)

(u'some', 1)

(u'how', 1)

(u'hello', 1)

(u'are', 1)

(u'to', 1)

(u'again', 1)

(u'you', 2)

(u'coffee', 1)

(u'lets', 1)

...

-------------------------------------------

Time: 2017-10-12 21:08:40

-------------------------------------------

(u'go', 1)

(u'some', 1)

(u'are', 1)

(u'how', 1)

(u'hello', 1)

(u'to', 1)

(u'again', 1)

(u'you', 2)

(u'coffee', 1)

(u'lets', 1)

...

-------------------------------------------

Time: 2017-10-12 21:08:41

-------------------------------------------

(u'go', 1)

(u'some', 1)

(u'how', 1)

(u'hello', 1)

(u'are', 1)

(u'to', 1)

(u'again', 1)

(u'you', 2)

(u'coffee', 1)

(u'lets', 1)

...

-------------------------------------------

Time: 2017-10-12 21:08:42

-------------------------------------------

(u'go', 1)

(u'some', 1)

(u'are', 1)

(u'how', 1)

(u'hello', 1)

(u'to', 1)

(u'again', 1)

(u'you', 2)

(u'coffee', 1)

(u'lets', 1)

...

-------------------------------------------

Time: 2017-10-12 21:08:43

-------------------------------------------

(u'go', 1)

(u'some', 1)

(u'how', 1)

(u'hello', 1)

(u'are', 1)

(u'to', 1)

(u'again', 1)

(u'you', 2)

(u'coffee', 1)

(u'lets', 1)

...

-------------------------------------------

Time: 2017-10-12 21:08:44

-------------------------------------------

(u'go', 1)

(u'some', 1)

(u'are', 1)

(u'how', 1)

(u'hello', 1)

(u'to', 1)

(u'again', 1)

(u'you', 2)

(u'coffee', 1)

(u'lets', 1)

...

-------------------------------------------

Time: 2017-10-12 21:08:45

-------------------------------------------

(u'go', 1)

(u'some', 1)

(u'how', 1)

(u'hello', 1)

(u'are', 1)

(u'to', 1)

(u'again', 1)

(u'you', 2)

(u'coffee', 1)

(u'lets', 1)

...

-------------------------------------------

Time: 2017-10-12 21:08:46

-------------------------------------------

(u'go', 1)

(u'some', 1)

(u'are', 1)

(u'how', 1)

(u'hello', 1)

(u'to', 1)

(u'again', 1)

(u'you', 2)

(u'coffee', 1)

(u'lets', 1)

...

-------------------------------------------

Time: 2017-10-12 21:08:47

-------------------------------------------

(u'go', 1)

(u'some', 1)

(u'how', 1)

(u'hello', 1)

(u'are', 1)

(u'to', 1)

(u'again', 1)

(u'you', 2)

(u'coffee', 1)

(u'lets', 1)

...

(b)*modified to only show all words with a or b*

from \_\_future\_\_ import print\_function

import sys

from pyspark import SparkContext

from pyspark.streaming import StreamingContext

if \_\_name\_\_ == "\_\_main\_\_":

if len(sys.argv) != 3:

print("Usage: stateful\_network\_wordcount.py <hostname> <port>", file=sys.stderr)

exit(-1)

sc = SparkContext(appName="PythonStreamingStatefulNetworkWordCount")

ssc = StreamingContext(sc, 1)

ssc.checkpoint("checkpoint")

# RDD with initial state (key, value) pairs

# initialStateRDD = sc.parallelize([(u'hello', 1), (u'world', 1)])

def updateFunc(new\_values, last\_sum):

return sum(new\_values) + (last\_sum or 0)

lines = ssc.socketTextStream(sys.argv[1], int(sys.argv[2]))

running\_counts = lines.flatMap(lambda line: line.split(" "))\

.filter(lambda wordWith:\

wordWith.startswith("a") or \

wordWith.startswith("b"))\

.map(lambda word: (word, 1))\

.updateStateByKey(updateFunc)

#.updateStateByKey(updateFunc,initialStateRDD)

running\_counts.pprint()

ssc.start()

ssc.awaitTermination**()**

*Results from running the above*

[rsha@localhost input]$ nc -lk 9999

I am

will be lost

i am

modern

classic

alpha

beta

Alpha

this will not be printed

apple will be counted

banana will be counted

orange will be skipped

[rsha@localhost hw6]$ spark-submit --master local[4] stateful\_wordcount.py localhost 9999

-------------------------------------------

Time: 2017-10-12 21:33:12

-------------------------------------------

-------------------------------------------

Time: 2017-10-12 21:33:13

-------------------------------------------

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:14

-------------------------------------------

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:15

-------------------------------------------

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:16

-------------------------------------------

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:17

-------------------------------------------

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:18

-------------------------------------------

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:19

-------------------------------------------

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:20

-------------------------------------------

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:21

-------------------------------------------

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:22

-------------------------------------------

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:23

-------------------------------------------

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:24

-------------------------------------------

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:25

-------------------------------------------

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:26

-------------------------------------------

(u'am', 1)

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:27

-------------------------------------------

(u'am', 1)

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:28

-------------------------------------------

(u'am', 1)

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:29

-------------------------------------------

(u'am', 1)

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:30

-------------------------------------------

(u'am', 1)

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:31

-------------------------------------------

(u'am', 1)

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:32

-------------------------------------------

(u'am', 1)

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:33

-------------------------------------------

(u'am', 1)

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:34

-------------------------------------------

(u'am', 1)

(u'be', 1)

-------------------------------------------

Time: 2017-10-12 21:33:35

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

-------------------------------------------

Time: 2017-10-12 21:33:36

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:37

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:38

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:39

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:40

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:41

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:42

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:43

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:44

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:45

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:46

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:47

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:48

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:49

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:50

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:51

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:52

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:53

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:54

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:55

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:56

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:57

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:58

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:33:59

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:00

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:01

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:02

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:03

-------------------------------------------

(u'am', 1)

(u'be', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:04

-------------------------------------------

(u'am', 1)

(u'be', 2)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:05

-------------------------------------------

(u'am', 1)

(u'be', 2)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:06

-------------------------------------------

(u'am', 1)

(u'be', 2)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:07

-------------------------------------------

(u'am', 1)

(u'be', 2)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:08

-------------------------------------------

(u'am', 1)

(u'be', 2)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:09

-------------------------------------------

(u'am', 1)

(u'be', 3)

(u'apple', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:10

-------------------------------------------

(u'am', 1)

(u'be', 3)

(u'apple', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:11

-------------------------------------------

(u'am', 1)

(u'be', 3)

(u'apple', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:12

-------------------------------------------

(u'am', 1)

(u'be', 3)

(u'apple', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:13

-------------------------------------------

(u'am', 1)

(u'be', 3)

(u'apple', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:14

-------------------------------------------

(u'am', 1)

(u'be', 3)

(u'apple', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:15

-------------------------------------------

(u'am', 1)

(u'be', 3)

(u'apple', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:16

-------------------------------------------

(u'am', 1)

(u'be', 3)

(u'apple', 1)

(u'alpha', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:17

-------------------------------------------

(u'am', 1)

(u'be', 4)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:18

-------------------------------------------

(u'am', 1)

(u'be', 4)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:19

-------------------------------------------

(u'am', 1)

(u'be', 4)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:20

-------------------------------------------

(u'am', 1)

(u'be', 4)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:21

-------------------------------------------

(u'am', 1)

(u'be', 4)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:22

-------------------------------------------

(u'am', 1)

(u'be', 4)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:23

-------------------------------------------

(u'am', 1)

(u'be', 4)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:24

-------------------------------------------

(u'am', 1)

(u'be', 4)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:25

-------------------------------------------

(u'am', 1)

(u'be', 4)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:26

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:27

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:28

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:29

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:30

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:31

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:32

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:33

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:34

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:35

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:36

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:37

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:38

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:39

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:40

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:41

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:42

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:43

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:44

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:45

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:46

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)

-------------------------------------------

Time: 2017-10-12 21:34:47

-------------------------------------------

(u'am', 1)

(u'be', 5)

(u'alpha', 1)

(u'apple', 1)

(u'banana', 1)

(u'beta', 1)