Q 6. WordCount2.py results:

a_to_n = 46

other = 49

```
[hadoop@ip-172-31-11-25 ~]$ python WordCount2.py -r hadoop hdfs:///user/hadoop/w.data
No configs found; falling back on auto-configuration
No configs specified for hadoop runner
Looking for hadoop binary in $PATH...
Found hadoop binary: /usr/bin/hadoop
Using Hadoop version 2.8.5
Looking for Hadoop streaming jar in /home/hadoop/contrib...
Looking for Hadoop streaming jar in /usr/lib/hadoop-mapreduce...
Found Hadoop streaming jar: /usr/lib/hadoop-mapreduce/hadoop-streaming.jar
Creating temp directory /tmp/WordCount2.hadoop.20200917.040634.005672
uploading working dir files to hdfs:///user/hadoop/tmp/mrjob/WordCount2.hadoop.20200917.040634.005672/files/wd...
Copying other local files to hdfs:///user/hadoop/tmp/mrjob/WordCount2.hadoop.20200917.040634.005672/files/
Running step 1 of 1...
  packageJobJar: [] [/usr/lib/hadoop/hadoop-streaming-2.8.5-amzn-6.jar] /tmp/streamjob7614291077697093352.jar tmpDir=null
  Connecting to ResourceManager at ip-172-31-11-25.ec2.internal/172.31.11.25:8032
  Connecting to ResourceManager at ip-172-31-11-25.ec2.internal/172.31.11.25:8032
  Loaded native gpl library
  Successfully loaded & initialized native-lzo library [hadoop-lzo rev ff8f5709577defb6b78cdc1f98cfe129c4b6fe46]
  Total input files to process : 1
  number of splits:4
  Submitting tokens for job: job_1600309860566_0005
  Submitted application application_1600309860566_0005
  The url to track the job: http://ip-172-31-11-25.ec2.internal:20888/proxy/application_1600309860566_0005/
  Running job: job_1600309860566_0005
  Job job_1600309860566_0005 running in uber mode : false
   map 0% reduce 0%
   map 50% reduce 0%
   map 75% reduce 0%
   map 100% reduce 0%
   map 100% reduce 100%
  Job job_1600309860566_0005 completed successfully
  Output directory: hdfs:///user/hadoop/tmp/mrjob/WordCount2.hadoop.20200917.040634.005672/output
Counters: 49
        File Input Format Counters
                Bytes Read=1320
        File Output Format Counters
                Bytes Written=23
        File System Counters
                FILE: Number of bytes read=78
                FILE: Number of bytes written=872372
                FILE: Number of large read operations=0
                FILE: Number of read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1764
                HDFS: Number of bytes written=23
                HDFS: Number of large read operations=0
                HDFS: Number of read operations=15
                HDFS: Number of write operations=2
        Job Counters
                Data-local map tasks=4
                Launched map tasks=4
                Launched reduce tasks=1
                Total megabyte-milliseconds taken by all map tasks=47542272
                Total megabyte-milliseconds taken by all reduce tasks=12020736
                Total time spent by all map tasks (ms)=30952
                Total time spent by all maps in occupied slots (ms)=1485696
                Total time spent by all reduce tasks (ms)=3913
                Total time spent by all reduces in occupied slots (ms)=375648
                Total vcore-milliseconds taken by all map tasks=30952
                Total vcore-milliseconds taken by all reduce tasks=3913
        Map-Reduce Framework
                CPU time spent (ms)=5330
                Combine input records=95
                Combine output records=6
                Failed Shuffles=0
                GC time elapsed (ms)=635
                Input split bytes=444
                Map input records=6
                Map output bytes=996
                Map output materialized bytes=144
                Map output records=95
                Merged Map outputs=4
                Physical memory (bytes) snapshot=1997877248
                Reduce input groups=2
                Reduce input records=6
                Reduce output records=2
                Reduce shuffle bytes=144
                Shuffled Maps =4
                Spilled Records=12
                Total committed heap usage (bytes)=1793064960
                Virtual memory (bytes) snapshot=17769648128
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
job output is in hdfs:///user/hadoop/tmp/mrjob/WordCount2.hadoop.20200917.040634.005672/output
Streaming final output from hdfs:///user/hadoop/tmp/mrjob/WordCount2.hadoop.20200917.040634.005672/output...
"a_to_n"        46
"other" 49
Removing HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/WordCount2.hadoop.20200917.040634.005672...
Removing temp directory /tmp/WordCount2.hadoop.20200917.040634.005672...
[hadoop@ip-172-31-11-25 ~]$
```

Q 10. Salaries2.py results:

High = 442

Low = 7064

Medium = 6312

```
[hadoop@ip-172-31-11-25 ~]$ python Salaries2.py -r hadoop hdfs:///user/hadoop/Salaries.tsv
No configs found; falling back on auto-configuration
No configs specified for hadoop runner
Looking for hadoop binary in $PATH...
Found hadoop binary: /usr/bin/hadoop
Using Hadoop version 2.8.5
Looking for Hadoop streaming jar in /home/hadoop/contrib...
Looking for Hadoop streaming jar in /usr/lib/hadoop-mapreduce...
Found Hadoop streaming jar: /usr/lib/hadoop-mapreduce/hadoop-streaming.jar
Creating temp directory /tmp/Salaries2.hadoop.20200917.043415.335771
uploading working dir files to hdfs:///user/hadoop/tmp/mrjob/Salaries2.hadoop.20200917.043415.335771/files/wd...
Copying other local files to hdfs:///user/hadoop/tmp/mrjob/Salaries2.hadoop.20200917.043415.335771/files/
Running step 1 of 1...
  packageJobJar: [] [/usr/lib/hadoop/hadoop-streaming-2.8.5-amzn-6.jar] /tmp/streamjob4111262436435661 8.jar tmpDir=null
  Connecting to ResourceManager at ip-172-31-11-25.ec2.internal/172.31.11.25:8032
  Connecting to ResourceManager at ip-172-31-11-25.ec2.internal/172.31.11.25:8032
  Loaded native gpl library
  Successfully loaded & initialized native-lzo library [hadoop-lzo rev ff8f5709577defb6b78cdc1f98cfe129c4b6fe46]
  Total input files to process : 1
  number of splits:4
  Submitting tokens for job: job_1600309860566_0007
  Submitted application application_1600309860566_0007
  The url to track the job: http://ip-172-31-11-25.ec2.internal:20888/proxy/application_1600309860566_0007/
  Running job: job_1600309860566_0007
  Job job_1600309860566_0007 running in uber mode : false
   map 0% reduce 0%
   map 50% reduce 0%
   map 75% reduce 0%
   map 100% reduce 0%
   map 100% reduce 100%
  Job job_1600309860566_0007 completed successfully
  Output directory: hdfs:///user/hadoop/tmp/mrjob/Salaries2.hadoop.20200917.043415.335771/output
Counters: 50
        File Input Format Counters
                Bytes Read=1564110
        File Output Format Counters
                Bytes Written=36
        File System Counters
                FILE: Number of bytes read=116
                FILE: Number of bytes written=872442
                FILE: Number of large read operations=0
                FILE: Number of read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1564578
                HDFS: Number of bytes written=36
                HDFS: Number of large read operations=0
                HDFS: Number of read operations=15
                HDFS: Number of write operations=2
        Job Counters
                Data-local map tasks=4
                Killed map tasks=1
                Launched map tasks=4
                Launched reduce tasks=1
                Total megabyte-milliseconds taken by all map tasks=50045952
                Total megabyte-milliseconds taken by all reduce tasks=11986944
                Total time spent by all map tasks (ms)=32582
                Total time spent by all maps in occupied slots (ms)=1563936
                Total time spent by all reduce tasks (ms)=3902
                Total time spent by all reduces in occupied slots (ms)=374592
                Total vcore-milliseconds taken by all map tasks=32582
                Total vcore-milliseconds taken by all reduce tasks=3902
        Map-Reduce Framework
                CPU time spent (ms)=6160
                Combine input records=13818
                Combine output records=12
                Failed Shuffles=0
                GC time elapsed (ms)=758
                Input split bytes=468
                Map input records=13818
                Map output bytes=129922
                Map output materialized bytes=231
                Map output records=13818
                Merged Map outputs=4
                Physical memory (bytes) snapshot=1960869888
                Reduce input groups=3
                Reduce input records=12
                Reduce output records=3
                Reduce shuffle bytes=231
                Shuffled Maps =4
                Spilled Records=24
                Total committed heap usage (bytes)=1756889088
                Virtual memory (bytes) snapshot=17766526976
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
job output is in hdfs:///user/hadoop/tmp/mrjob/Salaries2.hadoop.20200917.043415.335771/output
Streaming final output from hdfs:///user/hadoop/tmp/mrjob/Salaries2.hadoop.20200917.043415.335771/output...
"High"  442
"Low"   7064
"Medium"        6312
Removing HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/Salaries2.hadoop.20200917.043415.335771...
Removing temp directory /tmp/Salaries2.hadoop.20200917.043415.335771...
[hadoop@ip-172-31-11-25 ~]$
```

Q 12. Movies.py results:

```
[hadoop@ip-172-31-11-25 ~]$ python Movies.py -r hadoop hdfs:///user/hadoop/u.data
No configs found; falling back on auto-configuration
No configs specified for hadoop runner
Looking for hadoop binary in $PATH...
Found hadoop binary: /usr/bin/hadoop
Using Hadoop version 2.8.5
Looking for Hadoop streaming jar in /home/hadoop/contrib...
Looking for Hadoop streaming jar in /usr/lib/hadoop-mapreduce...
Found Hadoop streaming jar: /usr/lib/hadoop-mapreduce/hadoop-streaming.jar
Creating temp directory /tmp/Movies.hadoop.20200917.051443.931910
uploading working dir files to hdfs:///user/hadoop/tmp/mrjob/Movies.hadoop.20200917.051443.931910/files/wd...
Copying other local files to hdfs:///user/hadoop/tmp/mrjob/Movies.hadoop.20200917.051443.931910/files/
Running step 1 of 1...
  packageJobJar: [] [/usr/lib/hadoop/hadoop-streaming-2.8.5-amzn-6.jar] /tmp/streamjob1511228572715409823.jar tmpDir=null
  Connecting to ResourceManager at ip-172-31-11-25.ec2.internal/172.31.11.25:8032
  Connecting to ResourceManager at ip-172-31-11-25.ec2.internal/172.31.11.25:8032
  Loaded native gpl library
  Successfully loaded & initialized native-lzo library [hadoop-lzo rev ff8f5709577defb6b78cdc1f98cfe129c4b6fe46]
  Total input files to process : 1
  number of splits:4
  Submitting tokens for job: job_1600309860566_0011
  Submitted application application_1600309860566_0011
  The url to track the job: http://ip-172-31-11-25.ec2.internal:20888/proxy/application_1600309860566_0011/
  Running job: job_1600309860566_0011
  Job job_1600309860566_0011 running in uber mode : false
   map 0% reduce 0%
   map 50% reduce 0%
   map 75% reduce 0%
   map 100% reduce 0%
   map 100% reduce 100%
  Job job_1600309860566_0011 completed successfully
  Output directory: hdfs:///user/hadoop/tmp/mrjob/Movies.hadoop.20200917.051443.931910/output
Counters: 50
        File Input Format Counters
                Bytes Read=2575317
        File Output Format Counters
                Bytes Written=6204
        File System Counters
                FILE: Number of bytes read=4636
                FILE: Number of bytes written=881442
                FILE: Number of large read operations=0
                FILE: Number of read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=2575761
                HDFS: Number of bytes written=6204
                HDFS: Number of large read operations=0
                HDFS: Number of read operations=15
                HDFS: Number of write operations=2
        Job Counters
                Data-local map tasks=4
                Killed map tasks=1
                Launched map tasks=4
                Launched reduce tasks=1
                Total megabyte-milliseconds taken by all map tasks=53612544
                Total megabyte-milliseconds taken by all reduce tasks=12100608
                Total time spent by all map tasks (ms)=34904
                Total time spent by all maps in occupied slots (ms)=1675392
                Total time spent by all reduce tasks (ms)=3939
                Total time spent by all reduces in occupied slots (ms)=378144
                Total time spent by all maps in occupied slots (ms)=1675392
                Total time spent by all reduce tasks (ms)=3939
                Total time spent by all reduces in occupied slots (ms)=378144
                Total vcore-milliseconds taken by all map tasks=34904
                Total vcore-milliseconds taken by all reduce tasks=3939
        Map-Reduce Framework
                CPU time spent (ms)=6980
                Combine input records=100004
                Combine output records=674
                Failed Shuffles=0
                GC time elapsed (ms)=823
                Input split bytes=444
                Map input records=100004
                Map output bytes=784015
                Map output materialized bytes=4956
                Map output records=100004
                Merged Map outputs=4
                Physical memory (bytes) snapshot=1880678400
                Reduce input groups=671
                Reduce input records=674
                Reduce output records=671
                Reduce shuffle bytes=4956
                Shuffled Maps =4
                Spilled Records=1348
                Total committed heap usage (bytes)=1690828800
                Virtual memory (bytes) snapshot=17708642304
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
job output is in hdfs:///user/hadoop/tmp/mrjob/Movies.hadoop.20200917.051443.931910/output
Streaming final output from hdfs:///user/hadoop/tmp/mrjob/Movies.hadoop.20200917.051443.931910/output...
"1"     20
"10"    46
"100"   25
"101"   55
"102"   678
"103"   94
"104"   76
"105"   525
"106"   45
"107"   32
"108"   31
"109"   23
"11"    38
"110"   120
"111"   341
"112"   21
"113"   27
"114"   25
"115"   41
"116"   25
"117"   55
"118"   189
"119"   641
"12"    61
```