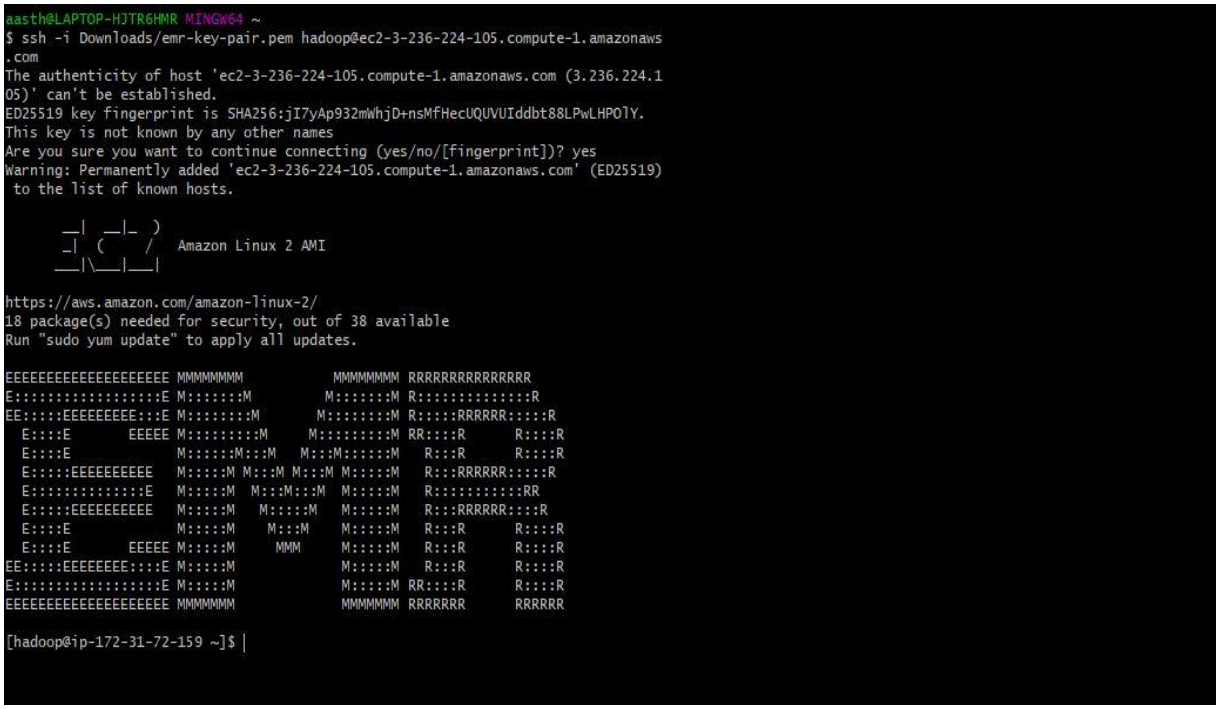
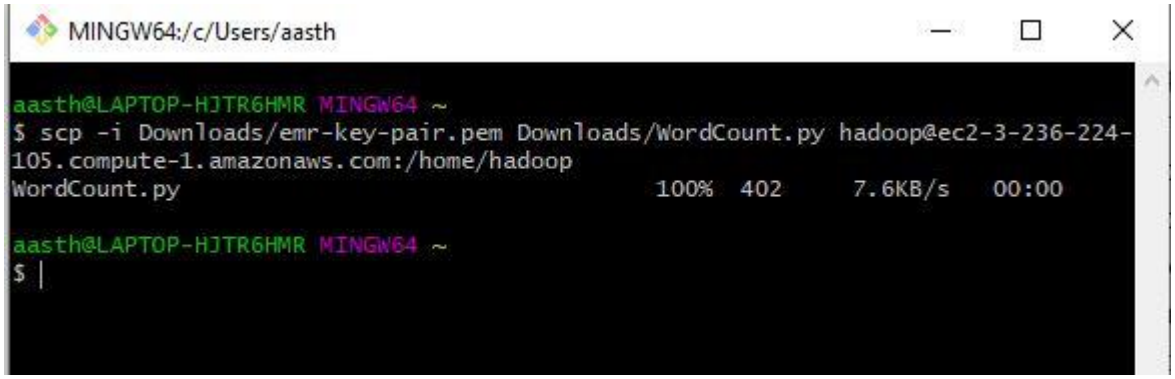
Big Data Technologies (CSP 554)

Assignment-3

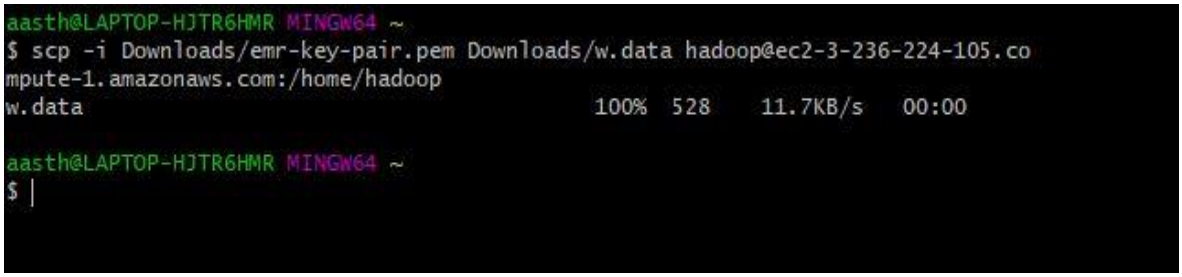
1. Doing ssh to the master node



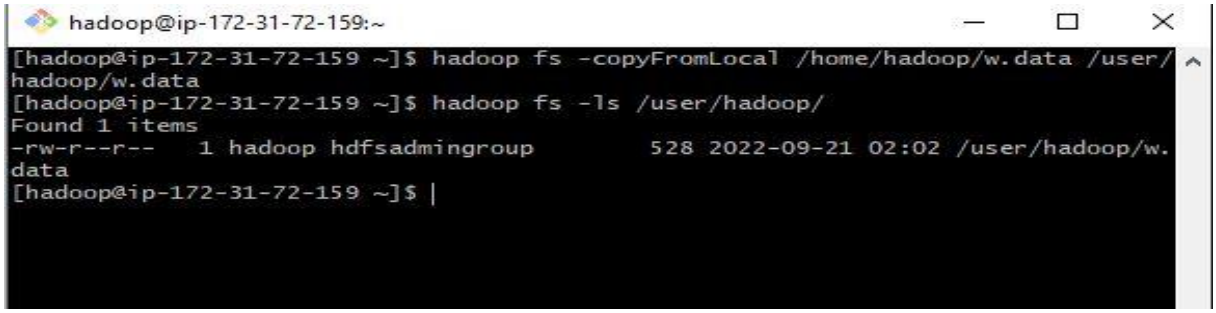
2. Moving WordCount.py to home/hadoop



3. Moving w.data to home/Hadoop



4. Moving w.data to user/Hadoop



5. Execute WordCount.py

```
[hadoop@ip-172-31-72-159 ~]$ python WordCount.py -r hadoop hdfs:///user/hadoop/w
.data --output-dir /user/hadoop/dout
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 2.10.1
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/WordCount.hadoop.20220921.020636.646864
uploading working dir files to hdfs:///user/hadoop/tmp/mrjob/WordCount.hadoop.20
220921.020636.646864/files/wd...
Copying other local files to hdfs:///user/hadoop/tmp/mrjob/WordCount.hadoop.2022
0921.020636.646864/files/
Running step 1 of 1...
  packageJobJar: [] [/usr/lib/hadoop/hadoop-streaming-2.10.1-amzn-4.jar] /tmp/st
reamjob1746678906089515848.jar tmpDir=null
  Connecting to ResourceManager at ip-172-31-72-159.ec2.internal/172.31.72.159:8
032
  Connecting to Application History server at ip-172-31-72-159.ec2.internal/172.
31.72.159:10200
  Connecting to ResourceManager at ip-172-31-72-159.ec2.internal/172.31.72.159:8
032
  Connecting to Application History server at ip-172-31-72-159.ec2.internal/172.
31.72.159:10200
  Loaded native gpl library
  Successfully loaded & initialized native-lzo library [hadoop-lzo rev 049362b7c
f53ff5f739d6b1532457f2c6cd495e8]
  Total input files to process : 1
  number of splits:4
  Submitting tokens for job: job_1663725352617_0001
  resource-types.xml not found
  Unable to find 'resource-types.xml'.
  Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE
  Adding resource type - name = vcores, units = , type = COUNTABLE
  Submitted application application_1663725352617_0001
  The url to track the job: http://ip-172-31-72-159.ec2.internal:20888/proxy/app
lication_1663725352617_0001/
  Running job: job_1663725352617_0001
  Job job_1663725352617_0001 running in uber mode : false
   map 0% reduce 0%
   map 50% reduce 0%
   map 100% reduce 0%
   map 100% reduce 100%
  Job job_1663725352617_0001 completed successfully
  Output directory: hdfs:///user/hadoop/dout
Counters: 50
        File Input Format Counters
                Bytes Read=1320
        File Output Format Counters
                Bytes Written=652
        File System Counters
                FILE: Number of bytes read=632
                FILE: Number of bytes written=1129422
                FILE: Number of large read operations=0
                FILE: Number of read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1768
                HDFS: Number of bytes written=652
                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=62191104
                Total megabyte-milliseconds taken by all reduce tasks=13135872
                Total time spent by all map tasks (ms)=40489
                Total time spent by all maps in occupied slots (ms)=1943472
                Total time spent by all reduce tasks (ms)=4276
                Total time spent by all reduces in occupied slots (ms)=410496
                Total vcore-milliseconds taken by all map tasks=40489
                Total vcore-milliseconds taken by all reduce tasks=4276
        Map-Reduce Framework
                CPU time spent (ms)=4620
                Combine input records=95
                Combine output records=80
                Failed Shuffles=0
                GC time elapsed (ms)=957
                Input split bytes=448
                Map input records=6
                Map output bytes=891
                Map output materialized bytes=805
                Map output records=95
                Merged Map outputs=4
                Physical memory (bytes) snapshot=2036060160
                Reduce input groups=65
                Reduce input records=80
                Reduce output records=65
                Reduce shuffle bytes=805
                Shuffled Maps =4
                Spilled Records=160
                Total committed heap usage (bytes)=1612185600
                Virtual memory (bytes) snapshot=17856229376
        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/dout
Removing HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/WordCount.hadoop.2022
0921.020636.646864...
Removing temp directory /tmp/WordCount.hadoop.20220921.020636.646864...
[hadoop@ip-172-31-72-159 ~]$
```

6. Output of WordCount.py

```
[hadoop@ip-172-31-72-159 ~]$ hadoop fs -cat /user/hadoop/WordCount/part-00000
"a"     3
"all"   1
"an"    1
"and"   1
"are"   1
"as"    4
"available"     1
"be"    3
"by"    1
"cluster"       2
"combine"       1
"contained"     1
"defined"       1
"dependencies"  1
"do"    1
"either"        1
"executed"      1
"explains"      1
"file"  2
"first" 1
"following"     1
"for"   1
"hadoop"        1
"how"   2
"in"    1
"individual"    1
"is"    2
"job"   4
"machine"       1
"map"   1
"more"  2
"mrjob" 1
"must"  1
"nodes" 1
"of"    1
"on"    4
```

6. Now slightly modify the WordCount.py program. Call the new program WordCount2.py.

Instead of counting how many words there are in the input documents (w.data), modify the program to count how many words begin with the small letters a-n and how many begin with anything else.

The output file should look something like

a_to_n, 12

other, 21

Now execute the program and see what happens.

6) (5 points) Submit a copy of this modified program and a screen shot of the results of the program's execution as the output of your assignment.

WordCount2.py Code

```
from mrjob.job import MRJob
import re

WORD_RE = re.compile(r"[\w']+")


class MRWordCount(MRJob):

    def mapper(self, _, line):
        for word in WORD_RE.findall(line):
            if re.match(r'[a-n]', word[0]):
                yield 'a_to_n', 1
            else:
                yield 'other', 1


    def combiner(self, word, counts):
        yield word, sum(counts)

    def reducer(self, word, counts):
        yield word, sum(counts)


if __name__ == '__main__':
    MRWordCount.run()
~
~
~
~
```

Execution of WordCount2.py

```
[hadoop@ip-172-31-72-159 ~]$ python WordCount2.py -r hadoop hdfs:///user/hadoop/
w.data --output-dir /user/hadoop/WordCount2
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 2.10.1
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.20220921.021917.258622
uploading working dir files to hdfs:///user/hadoop/tmp/mrjob/WordCount2.hadoop.2
0220921.021917.258622/files/wd...
Copying other local files to hdfs:///user/hadoop/tmp/mrjob/WordCount2.hadoop.202
20921.021917.258622/files/
Running step 1 of 1...
  packageJobJar: [] [/usr/lib/hadoop/hadoop-streaming-2.10.1-amzn-4.jar] /tmp/st
reamjob7509234584707718635.jar tmpDir=null
  Connecting to ResourceManager at ip-172-31-72-159.ec2.internal/172.31.72.159:8
032
  Connecting to Application History server at ip-172-31-72-159.ec2.internal/172.
31.72.159:10200
  Connecting to ResourceManager at ip-172-31-72-159.ec2.internal/172.31.72.159:8
032
  Connecting to Application History server at ip-172-31-72-159.ec2.internal/172.
31.72.159:10200
  Loaded native gpl library
  Successfully loaded & initialized native-lzo library [hadoop-lzo rev 049362b7c
f53ff5f739d6b1532457f2c6cd495e8]
  Total input files to process : 1
  number of splits:4
  Submitting tokens for job: job_1663725352617_0002
  resource-types.xml not found
  Unable to find 'resource-types.xml'.
  Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE
  Adding resource type - name = vcores, units = , type = COUNTABLE
  Submitted application application_1663725352617_0002
  The url to track the job: http://ip-172-31-72-159.ec2.internal:20888/proxy/app
lication_1663725352617_0002/
  Running job: job_1663725352617_0002
  Job job_1663725352617_0002 running in uber mode : false
   map 0% reduce 0%
   map 25% reduce 0%
   map 50% reduce 0%
   map 75% reduce 0%
   map 100% reduce 0%
   map 100% reduce 100%
  Job job_1663725352617_0002 completed successfully
  Output directory: hdfs:///user/hadoop/WordCount2
Counters: 50
        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=1128307
                FILE: Number of large read operations=0
                FILE: Number of read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1768
                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
```

Output of WordCount2.py

```
[hadoop@ip-172-31-72-159 ~]$ hadoop fs -cat /user/hadoop/WordCount2/part-00000
"a_to_n"        46
"other" 49
[hadoop@ip-172-31-72-159 ~]$
```

8) Execute the Salaries.py program to make sure it works. It should print out how many workers share each job title.

9) Now modify the Salaries.py program. Call it Salaries2.py

Instead of counting the number of workers per department, change the program to provide the number of workers having High, Medium or Low annual salaries. This is defined as follows:

High	**100,000.00 and above**
Medium	**50,000.00 to 99,999.99**
Low	**0.00 to 49,999.99**

The output of the program should be something like the following (in any order):

High 20

Medium 30

Low 10

10) (5 points) Submit a copy of this modified program and a screenshot of the results of the program's execution as the output of your assignment.

Salaries2.py Code

```
from mrjob.job import MRJob

class MRSalaries(MRJob):

    def mapper(self, _, line):
        (name,jobTitle,agencyID,agency,hireDate,annualSalary,grossPay) = line.split('\t')
        if(float(annualSalary) >= 100000.00):
            yield 'High', 1
        elif(float(annualSalary) >= 50000.0 and float(annualSalary)<=99999.99):
            yield 'Medium',1
        elif(float(annualSalary) >= 0.0 and float(annualSalary) <= 49999.99):
            yield 'Low', 1

    def combiner(self, jobTitle, counts):
        yield jobTitle, sum(counts)

    def reducer(self, jobTitle, counts):
        yield jobTitle, sum(counts)


if __name__ == '__main__':
    MRSalaries.run()
~
~
~
~
~
```

Executing Salaries2.py

```
[hadoop@ip-172-31-72-159 ~]$ vim Salaries2.py
[hadoop@ip-172-31-72-159 ~]$ python Salaries2.py -r hadoop hdfs:///user/hadoop/S
alaries.tsv --output-dir /user/hadoop/Salaries2
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 2.10.1
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.20220921.030716.293124
uploading working dir files to hdfs:///user/hadoop/tmp/mrjob/Salaries2.hadoop.20
220921.030716.293124/files/wd...
Copying other local files to hdfs:///user/hadoop/tmp/mrjob/Salaries2.hadoop.2022
0921.030716.293124/files/
Running step 1 of 1...
  packageJobJar: [] [/usr/lib/hadoop/hadoop-streaming-2.10.1-amzn-4.jar] /tmp/st
reamjob1039612061511943751.jar tmpDir=null
  Connecting to ResourceManager at ip-172-31-72-159.ec2.internal/172.31.72.159:8
032
  Connecting to Application History server at ip-172-31-72-159.ec2.internal/172.
31.72.159:10200
  Connecting to ResourceManager at ip-172-31-72-159.ec2.internal/172.31.72.159:8
032
  Connecting to Application History server at ip-172-31-72-159.ec2.internal/172.
31.72.159:10200
  Loaded native gpl library
  Successfully loaded & initialized native-lzo library [hadoop-lzo rev 049362b7c
f53ff5f739d6b1532457f2c6cd495e8]
  Total input files to process : 1
  number of splits:4
  Submitting tokens for job: job_1663725352617_0003
  resource-types.xml not found
  Unable to find 'resource-types.xml'.
  Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE
  Adding resource type - name = vcores, units = , type = COUNTABLE
  Submitted application application_1663725352617_0003
  The url to track the job: http://ip-172-31-72-159.ec2.internal:20888/proxy/app
lication_1663725352617_0003/
  Running job: job_1663725352617_0003
  Job job_1663725352617_0003 running in uber mode : false
   map 0% reduce 0%
   map 50% reduce 0%
   map 100% reduce 0%
   map 100% reduce 100%
  Job job_1663725352617_0003 completed successfully
  Output directory: hdfs:///user/hadoop/Salaries2
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=1128387
                FILE: Number of large read operations=0
```

```
                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=62204928
                Total megabyte-milliseconds taken by all reduce tasks=12558336
                Total time spent by all map tasks (ms)=40498
                Total time spent by all maps in occupied slots (ms)=1943904
                Total time spent by all reduce tasks (ms)=4088
                Total time spent by all reduces in occupied slots (ms)=392448
                Total vcore-milliseconds taken by all map tasks=40498
                Total vcore-milliseconds taken by all reduce tasks=4088
        Map-Reduce Framework
                CPU time spent (ms)=5860
                Combine input records=13818
                Combine output records=12
                Failed Shuffles=0
                GC time elapsed (ms)=1063
                Input split bytes=472
                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=2070986752
                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)=1643118592
                Virtual memory (bytes) snapshot=17883275264
        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/Salaries2
Removing HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/Salaries2.hadoop.2022
0921.030716.293124...
Removing temp directory /tmp/Salaries2.hadoop.20220921.030716.293124...
[hadoop@ip-172-31-72-159 ~]$
```

The output of Salaries2.py

```
[hadoop@ip-172-31-72-159 ~]$ hadoop fs -cat /user/hadoop/Salaries2/part-00000
"High"  442
"Low"   7064
"Medium"        6312
[hadoop@ip-172-31-72-159 ~]$
```

11) Now copy the file u.data from the assignment to /user/hadoop. This is similar to the file used for some examples in Module 03b. NOTE: unlike the slide deck examples, this version of u.data has fields separated by commas and not tabs.

12) (5 points) Review slides 55-61 in lecture notes Module 3b. Now write a program to perform the task of outputting a count of the number of movies each user (identified via their user id) reviewed.

Output might look something like the following:

186: 2

192: 2

112: 1

etc.

Submit a copy of this program and a screenshot of the results of the program's execution (only 10 lines or so of the result) as the output of your assignment.

Movies_rating Code

hadoop@ip-172-31-72-159:~

```
from mrjob.job import MRJob

class MRMovies(MRJob):

    def mapper(self, _, line):
        (user_id, movie_id, rating, timeStamp) = line.split(',')
        yield user_id, 1

    def combiner(self, user_id, counts):
        yield user_id, sum(counts)

    def reducer(self, user_id, counts):
        yield user_id, sum(counts)


if __name__ == '__main__':
    MRMovies.run()
~
~
~
~
```

Execution of Movies_rating Code

```
[hadoop@ip-172-31-72-159 ~]$ vim Movies_rating.py
[hadoop@ip-172-31-72-159 ~]$ python Movies_rating.py -r hadoop hdfs:///user/hado
op/u.data --output-dir /user/hadoop/Movies_rating
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 2.10.1
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_rating.hadoop.20220921.031915.720802
uploading working dir files to hdfs:///user/hadoop/tmp/mrjob/Movies_rating.hadoo
p.20220921.031915.720802/files/wd...
Copying other local files to hdfs:///user/hadoop/tmp/mrjob/Movies_rating.hadoop.
20220921.031915.720802/files/
Running step 1 of 1...
  packageJobJar: [] [/usr/lib/hadoop/hadoop-streaming-2.10.1-amzn-4.jar] /tmp/st
reamjob7655544157616678562.jar tmpDir=null
  Connecting to ResourceManager at ip-172-31-72-159.ec2.internal/172.31.72.159:8
032
  Connecting to Application History server at ip-172-31-72-159.ec2.internal/172.
31.72.159:10200
  Connecting to ResourceManager at ip-172-31-72-159.ec2.internal/172.31.72.159:8
032
  Connecting to Application History server at ip-172-31-72-159.ec2.internal/172.
31.72.159:10200
  Loaded native gpl library
  Successfully loaded & initialized native-lzo library [hadoop-lzo rev 049362b7c
f53ff5f739d6b1532457f2c6cd495e8]
  Total input files to process : 1
  number of splits:4
  Submitting tokens for job: job_1663725352617_0004
  resource-types.xml not found
  Unable to find 'resource-types.xml'.
  Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE
  Adding resource type - name = vcores, units = , type = COUNTABLE
  Submitted application application_1663725352617_0004
  The url to track the job: http://ip-172-31-72-159.ec2.internal:20888/proxy/app
lication_1663725352617_0004/
  Running job: job_1663725352617_0004
  Job job_1663725352617_0004 running in uber mode : false
   map 0% reduce 0%
   map 25% reduce 0%
   map 50% reduce 0%
   map 100% reduce 0%
   map 100% reduce 100%
  Job job_1663725352617_0004 completed successfully
  Output directory: hdfs:///user/hadoop/Movies_rating
Counters: 50
        File Input Format Counters
                Bytes Read=2575317
```

```
lication_1663725352617_0004/
  Running job: job_1663725352617_0004
  Job job_1663725352617_0004 running in uber mode : false
   map 0% reduce 0%
   map 25% reduce 0%
   map 50% reduce 0%
   map 100% reduce 0%
   map 100% reduce 100%
  Job job_1663725352617_0004 completed successfully
  Output directory: hdfs:///user/hadoop/Movies_rating
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=1137902
                FILE: Number of large read operations=0
                FILE: Number of read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=2575765
                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=70146048
                Total megabyte-milliseconds taken by all reduce tasks=12985344
                Total time spent by all map tasks (ms)=45668
                Total time spent by all maps in occupied slots (ms)=2192064
                Total time spent by all reduce tasks (ms)=4227
                Total time spent by all reduces in occupied slots (ms)=405792
                Total vcore-milliseconds taken by all map tasks=45668
                Total vcore-milliseconds taken by all reduce tasks=4227
        Map-Reduce Framework
                CPU time spent (ms)=8040
                Combine input records=100004
                Combine output records=674
                Failed Shuffles=0
                GC time elapsed (ms)=935
                Input split bytes=448
                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=2029432832
                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)=1599602688
                Virtual memory (bytes) snapshot=17849294848
        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/Movies_rating
Removing HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/Movies_rating.hadoop.
20220921.031915.720802...
Removing temp directory /tmp/Movies_rating.hadoop.20220921.031915.720802...
[hadoop@ip-172-31-72-159 ~]$
```

Output of Movies_rating.py

```
Removing temp directory /tmp/Movies_rating.hadoop.20220921.031915.720802...
[hadoop@ip-172-31-72-159 ~]$ hadoop fs -cat /user/hadoop/Movies_rating/part-0000
0
"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
"120"   138
"121"   80
"122"   40
"123"   33
"124"   85
"125"   210
"126"   64
"127"   21
```

```
"64"    21
"640"   49
"641"   140
"642"   36
"643"   24
"644"   39
"645"   30
"646"   169
"647"   150
"648"   256
"649"   90
"65"    27
"650"   29
"651"   20
"652"   267
"653"   51
"654"   626
"655"   105
"656"   128
"657"   20
"658"   60
"659"   142
"66"    49
"660"   92
"661"   33
"662"   58
"663"   26
"664"   519
"665"   434
"666"   40
"667"   68
"668"   20
"669"   37
"67"    103
"670"   31
"671"   115
"68"    123
"69"    81
"7"     88
"70"    83
"71"    23
"72"    191
"73"    1610
"74"    49
"75"    145
"76"    20
"77"    315
"78"    263
"79"    55
"8"     116
"80"    37
"81"    160
"82"    39
"83"    161
"84"    116
"85"    107
"86"    190
"87"    31
"88"    255
"89"    66
"9"     45
"90"    50
"91"    150
"92"    123
"93"    159
"94"    196
"95"    299
"96"    76
"97"    128
"98"    71
"99"    188
[hadoop@ip-172-31-72-159 ~]$
```

Submitted By:-

Aastha Dhir

CWID-A20468022

adhir2@hawk.iit.edu