CSP 554 BIG DATA TECHNOLOGIES
MOHAMMED WASIM R D(A20497053)

ASSIGNMENT 3

Step 1: Create a EMR cluster and install mrjob library

```
Last login: Mon Sep 19 18:16:43 on ttys000
(base) mohammedwasimrd@dhcp197 ~ % chmod 400 /Users/mohammedwasimrd/Downloads/ec2keypair.pem
(base) mohammedwasimrd@dhcp197 ~ % ssh -i /Users/mohammedwasimrd/Downloads/ec2keypair.pem hadoop@ec2-18-191-243-167.us-east-2.compute.amazonaws.com

The authenticity of host 'ec2-18-191-243-167.us-east-2.compute.amazonaws.com (18.191.243.167)' can't be established.
ED25519 key fingerprint is SHA256:1Kv2YIsCv9TqmB7ebHQ5SLexljLpb4AilUj+K1qrNNI.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'ec2-18-191-243-167.us-east-2.compute.amazonaws.com' (ED25519) to the list of known hosts.
Last login: Mon Sep 19 23:18:46 2022

       __|  __|_  )
       _|  (     /   Amazon Linux 2 AMI
      ___|\___|___|

https://aws.amazon.com/amazon-linux-2/
18 package(s) needed for security, out of 38 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMMM           MMMMMMMM RRRRRRRRRRRRRRR
E::::::::::::::::::E M:::::::M         M:::::::M R::::::::::::::R
EE:::::EEEEEEEEE:::E M::::::::M       M::::::::M R:::::RRRRRR:::::R
  E::::E       EEEEE M:::::::::M     M:::::::::M RR::::R      R::::R
  E::::E             M::::::M:::M   M:::M::::::M   R:::R      R::::R
  E:::::EEEEEEEEEE   M:::::M M:::M M:::M M:::::M   R:::RRRRRR:::::R
  E:::::::::::::::E  M:::::M  M:::M:::M  M:::::M   R:::::::::::RR
  E::::::EEEEEEEEEE  M:::::M   M:::::M   M:::::M   R:::RRRRRR:::::R
  E::::E             M:::::M    M:::M    M:::::M   R:::R      R::::R
  E::::E       EEEEE M:::::M     MMM     M:::::M   R:::R      R::::R
EE::::::EEEEEEEE:::E M:::::M             M:::::M   R:::R      R::::R
E::::::::::::::::::E M:::::M             M:::::M RR::::R      R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMM             MMMMMM RRRRRRR      RRRRRR

[hadoop@ip-172-31-38-254 ~]$ sudo /usr/bin/pip3.7 install mrjob[aws]
WARNING: Running pip install with root privileges is generally not a good idea. Try `pip3.7 install --user` instead.
Collecting mrjob[aws]
  Downloading mrjob-0.7.4-py2.py3-none-any.whl (439 kB)
     |████████████████████████████████| 439 kB 5.8 MB/s
Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib64/python3.7/site-packages (from mrjob[aws]) (5.4.1)
Collecting boto3>=1.10.0; extra == "aws"
  Downloading boto3-1.24.76-py3-none-any.whl (132 kB)
     |████████████████████████████████| 132 kB 47.0 MB/s
Collecting botocore>=1.13.26; extra == "aws"
  Downloading botocore-1.27.76-py3-none-any.whl (9.1 MB)
     |████████████████████████████████| 9.1 MB 45.2 MB/s
Collecting s3transfer<0.7.0,>=0.6.0
  Downloading s3transfer-0.6.0-py3-none-any.whl (79 kB)
     |████████████████████████████████| 79 kB 13.9 MB/s
Requirement already satisfied: jmespath<2.0.0,>=0.7.1 in /usr/local/lib/python3.7/site-packages (from boto3>=1.10.0; extra == "aws"->mrjob[aws]) (1.0.0)
Collecting python-dateutil<3.0.0,>=2.1
  Downloading python_dateutil-2.8.2-py2.py3-none-any.whl (247 kB)
     |████████████████████████████████| 247 kB 45.7 MB/s
Collecting urllib3<1.27,>=1.25.4
  Downloading urllib3-1.26.12-py2.py3-none-any.whl (140 kB)
     |████████████████████████████████| 140 kB 45.1 MB/s
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/site-packages (from python-dateutil<3.0.0,>=2.1->botocore>=1.13.26; extra == "aws"->mrjob[aws]) (1.13.0)
Installing collected packages: python-dateutil, urllib3, botocore, s3transfer, boto3, mrjob
  WARNING: The scripts mrjob, mrjob-3 and mrjob-3.7 are installed in '/usr/local/bin' which is not on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-script-location.
Successfully installed boto3-1.24.76 botocore-1.27.76 mrjob-0.7.4 python-dateutil-2.8.2 s3transfer-0.6.0 urllib3-1.26.12
[hadoop@ip-172-31-38-254 ~]$
```

Step 2: Set up wordcount.py and w.data from local to Hadoop master node and put the file in Hadoop file system directory.

```
[hadoop@ip-172-31-38-254 ~]$ hadoop fs -put w.data /user/hadoop/
[hadoop@ip-172-31-38-254 ~]$ hadoop fs -ls
Found 1 items
-rw-r--r--   1 hadoop hdfsadmingroup        528 2022-09-19 23:44 w.data
[hadoop@ip-172-31-38-254 ~]$ hadoop fs -ls /user/hadoop
Found 1 items
-rw-r--r--   1 hadoop hdfsadmingroup        528 2022-09-19 23:44 /user/hadoop/w.data
[hadoop@ip-172-31-38-254 ~]$
```

Step 3 : Execute wordcount.py with w.data

```
"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
"or"    2
"oriented"      1
"our"   1
"program"       1
"python"        1
"reduce"        1
"reference"     1
"run"   1
"runners"       1
"script"        1
"second"        1
"sections"      1
"see"   1
"submitted"     1
"task"  2
"that"  1
"the"   4
"things"        1
"those" 1
"to"    3
"two"   1
"uploaded"      1
"versions"      1
"well"  1
"when"  1
"will"  1
"within"        1
"writing"       2
"your"  5
Removing HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/WordCount.hadoop.20220919.235138.211000...
Removing temp directory /tmp/WordCount.hadoop.20220919.235138.211000...
[hadoop@ip-172-31-38-254 ~]$
```

Step 4: Modify wordcount.py to wordcount2.py and execute

```python
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(word[0]>="a" and word[0]<="n"):
                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()

```

```
job output is in hdfs:///user/hadoop/tmp/mrjob/WordCount2.hadoop.20220920.002339.988895/output
Streaming final output from hdfs:///user/hadoop/tmp/mrjob/WordCount2.hadoop.20220920.002339.988895/output...
"a_to_n"        46
"other" 49
Removing HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/WordCount2.hadoop.20220920.002339.988895...
Removing temp directory /tmp/WordCount2.hadoop.20220920.002339.988895...
[hadoop@ip-172-31-38-254 ~]$
```

Step 5: Same process for Salaries.py and Salaries.tsv. Add those to hadoop directory from Local system. Then transfer to hdfs to the hadoop directory.

Execute salaries.py

```
        at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:177)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1926)
        at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:171)


(from line 61 of hdfs:///tmp/hadoop-yarn/staging/history/done_intermediate/hadoop/job_1663628638628_0007-1663634620327-hadoop-streamjob960470665276232724.jar-1663634679465-1-0-FAILED-default-1663634627055
.jhist)


Step 1 of 1 failed: Command '['/usr/bin/hadoop', 'jar', '/usr/lib/hadoop-mapreduce/hadoop-streaming.jar', '-files', 'hdfs:///user/hadoop/tmp/mrjob/Salaries.hadoop.20220920.004312.623922/files/wd/Salaries.
py#Salaries.py,hdfs:///user/hadoop/tmp/mrjob/Salaries.hadoop.20220920.004312.623922/files/wd/mrjob.zip#mrjob.zip,hdfs:///user/hadoop/tmp/mrjob/Salaries.hadoop.20220920.004312.623922/files/wd/setup-wrapper
.sh#setup-wrapper.sh', '-input', 'hdfs:///user/hadoop/w.data', '-output', 'hdfs:///user/hadoop/tmp/mrjob/Salaries.hadoop.20220920.004312.623922/output', '-mapper', '/bin/sh -ex setup-wrapper.sh python3 Sa
[hadoop@ip-172-31-38-254 ~]$ python Salaries.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 2.10.1
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/Salaries.hadoop.20220920.004958.886820
uploading working dir files to hdfs:///user/hadoop/tmp/mrjob/Salaries.hadoop.20220920.004958.886820/files/wd...
Copying other local files to hdfs:///user/hadoop/tmp/mrjob/Salaries.hadoop.20220920.004958.886820/files/
Running step 1 of 1...
  packageJobJar: [] [/usr/lib/hadoop/hadoop-streaming-2.10.1-amzn-4.jar] /tmp/streamjob7610924344867422260.jar tmpDir=null
  Connecting to ResourceManager at ip-172-31-38-254.us-east-2.compute.internal/172.31.38.254:8032
  Connecting to Application History server at ip-172-31-38-254.us-east-2.compute.internal/172.31.38.254:10200
  Connecting to ResourceManager at ip-172-31-38-254.us-east-2.compute.internal/172.31.38.254:8032
  Connecting to Application History server at ip-172-31-38-254.us-east-2.compute.internal/172.31.38.254:10200
  Loaded native gpl library
  Successfully loaded & initialized native-lzo library [hadoop-lzo rev 049362b7cf53ff5f739d6b1532457f2c6cd495e8]
  Total input files to process : 1
  number of splits:4
  Submitting tokens for job: job_1663628638628_0008
  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_1663628638628_0008
  The url to track the job: http://ip-172-31-38-254.us-east-2.compute.internal:20888/proxy/application_1663628638628_0008/
  Running job: job_1663628638628_0008
  Job job_1663628638628_0008 running in uber mode : false
   map 0% reduce 0%
   map 50% reduce 0%
   map 100% reduce 0%
   map 100% reduce 100%
  Job job_1663628638628_0008 completed successfully
  Output directory: hdfs:///user/hadoop/tmp/mrjob/Salaries.hadoop.20220920.004958.886820/output
Counters: 50
        File Input Format Counters
                Bytes Read=1564110
        File Output Format Counters
                Bytes Written=29260
        File System Counters
                FILE: Number of bytes read=21270
                FILE: Number of bytes written=1187944
                FILE: Number of large read operations=0
                FILE: Number of read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1564638
```

```
"UTILITIES INSTALLER REPAIR S I"        19
"UTILITIES INSTALLER REPAIR SII"        15
"UTILITY AIDE"  15
"UTILITY INVESTIGATOR SUPV"     3
"UTILITY INVESTIGATOR"  12
"UTILITY METER FIELD OPER MANAG"        1
"UTILITY METER READER I"        23
"UTILITY METER READER II"       12
"UTILITY METER READER SUPT II"  1
"UTILITY METER READER SUPV"     5
"UTILITY POLICY ANALYST"        1
"Urban Forester"        7
"VEHICLE IDENTIFICATION INSPECT"        1
"VEHICLE PROCESSOR"     9
"VICE PRESIDENT CITY COUNCIL"   1
"VICTIM/WITNESS COORDINATOR SAO"        11
"VOLUNTEER SERVICE COORDINATOR" 1
"VOLUNTEER SERVICE WORKER"      1
"Volunteer Service Coordinator" 1
"WASTE WATER PLANT COORDINATOR" 2
"WASTE WATER PLANT MANAGER"     2
"WASTE WATER PLANT OPNS SUPV"   2
"WATER PUMPING ASST MANAGER"    2
"WATER SERVICE INSPECTOR"       4
"WATER SERVICE REPRESENTATIVE"  12
"WATER TREATMENT ASST MANAGER"  2
"WATER TREATMENT TECHNICIAN II" 17
"WATER TREATMENT TECHNICIAN III"        8
"WATER TREATMENT TECHNICIAN SUP"        6
"WATERSHED MAINT SUPV"  3
"WATERSHED MANAGER"     1
"WATERSHED RANGER II"   5
"WATERSHED RANGER III"  3
"WATERSHED RANGER SUPERVISOR"   1
"WEB DEVELOPER" 1
"WELDER"        8
"WHITEPRINT MACHINE OPR"        1
"WORK STUDY STUDENT"    18
"WORKER'S COMPENSATION CONTRACT"        1
"WWW Chief of Engineering"      1
"WWW Division  Manager I"       1
"WWW Division Manager II"       5
"Waste Water Maint Mgr Instrum" 1
"Waste Water Maintenance Mgr Me"        1
"Waste Water Opns Tech II Pump" 10
"Waste Water Opns Tech II Sanit"        81
"Waste Water Tech Supv I Pump"  6
"Waste Water Tech Supv II Pump" 1
"Waste Water Tech Supv II Sanit"        10
"Waste Water Techn Supv I Sanit"        19
"Water Systems Pumping Supv"    1
"Water Systems Treatment Manage"        1
"Water Systems Treatment Supv"  2
"YOUTH DEVELOPMENT TECH"        3
"ZONING ADMINISTRATOR"  1
"ZONING APPEALS ADVISOR BMZA"   1
"ZONING APPEALS OFFICER"        1
"ZONING ENFORCEMENT OFFICER"    1
"ZONING EXAMINER I"     2
"ZONING EXAMINER II"    1
Removing HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/Salaries.hadoop.20220920.004958.886820...
Removing temp directory /tmp/Salaries.hadoop.20220920.004958.886820...
[hadoop@ip-172-31-38-254 ~]$
```

Step 6: Modify Salaries.py to Salaries2.py

In [ ]:

```python
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)>=0.0 and float(annualSalary)<=49999.99:
            yield "Low", 1
        elif float(annualSalary)>=50000.0 and float(annualSalary)<=99999.99:
            yield "Medium", 1
        elif float(annualSalary)>=100000.0:
            yield "High", 1
    def combiner(self, jobTitle, counts):
        yield jobTitle, sum(counts)

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


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

```
		Bytes Read=1564110
	File Output Format Counters
		Bytes Written=36
	File System Counters
		FILE: Number of bytes read=116
		FILE: Number of bytes written=1131807
		FILE: Number of large read operations=0
		FILE: Number of read operations=0
		FILE: Number of write operations=0
		HDFS: Number of bytes read=1564638
		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=64046592
		Total megabyte-milliseconds taken by all reduce tasks=12954624
		Total time spent by all map tasks (ms)=41697
		Total time spent by all maps in occupied slots (ms)=2001456
		Total time spent by all reduce tasks (ms)=4217
		Total time spent by all reduces in occupied slots (ms)=404832
		Total vcore-milliseconds taken by all map tasks=41697
		Total vcore-milliseconds taken by all reduce tasks=4217
	Map-Reduce Framework
		CPU time spent (ms)=7480
		Combine input records=13818
		Combine output records=12
		Failed Shuffles=0
		GC time elapsed (ms)=920
		Input split bytes=528
		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=2074116096
		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)=1685061632
		Virtual memory (bytes) snapshot=17874837504
	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.20220920.010651.954594/output
Streaming final output from hdfs:///user/hadoop/tmp/mrjob/Salaries2.hadoop.20220920.010651.954594/output...
"High"  442
"Low"   7064
"Medium"        6312
Removing HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/Salaries2.hadoop.20220920.010651.954594...
Removing temp directory /tmp/Salaries2.hadoop.20220920.010651.954594...
[hadoop@ip-172-31-38-254 ~]$
```

Step 7 : Transfer u.data from Local to hadoop.

```
[hadoop@ip-172-31-38-254 ~]$ hadoop fs -put u.data /user/hadoop/
[hadoop@ip-172-31-38-254 ~]$ hadoop fs -ls /user/hadoop
Found 6 items
-rw-r--r--   1 hadoop hdfsadmingroup        411 2022-09-20 00:33 /user/hadoop/Salaries.py
-rw-r--r--   1 hadoop hdfsadmingroup        711 2022-09-20 01:06 /user/hadoop/Salaries2.py
-rw-r--r--   1 hadoop hdfsadmingroup    1538148 2022-09-20 00:36 /user/hadoop/salaries.tsv
drwxr-xr-x   - hadoop hdfsadmingroup          0 2022-09-19 23:46 /user/hadoop/tmp
-rw-r--r--   1 hadoop hdfsadmingroup    2438233 2022-09-20 01:17 /user/hadoop/u.data
-rw-r--r--   1 hadoop hdfsadmingroup        528 2022-09-19 23:44 /user/hadoop/w.data
[hadoop@ip-172-31-38-254 ~]$
```

Step 8 : Python program to output the count of movies of each user.

In [ ]:
```
from mrjob.job import MRJob

class MRMovies(MRJob):

    def mapper(self, _, line):
        (userid,movieid,rating,timestamp)=line.split(',')
        yield userid, 1

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

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


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

```
[hadoop@ip-172-31-38-254 ~]$ 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 2.10.1
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.hadoop.20220920.014542.512809
uploading working dir files to hdfs:///user/hadoop/tmp/mrjob/Movies.hadoop.20220920.014542.512809/files/wd...
Copying other local files to hdfs:///user/hadoop/tmp/mrjob/Movies.hadoop.20220920.014542.512809/files/
Running step 1 of 1...
  packageJobJar: [] [/usr/lib/hadoop/hadoop-streaming-2.10.1-amzn-4.jar] /tmp/streamjob6855650507179691095.jar tmpDir=null
  Connecting to ResourceManager at ip-172-31-38-254.us-east-2.compute.internal/172.31.38.254:8032
  Connecting to Application History server at ip-172-31-38-254.us-east-2.compute.internal/172.31.38.254:10200
  Connecting to ResourceManager at ip-172-31-38-254.us-east-2.compute.internal/172.31.38.254:8032
  Connecting to Application History server at ip-172-31-38-254.us-east-2.compute.internal/172.31.38.254:10200
  Loaded native gpl library
  Successfully loaded & initialized native-lzo library [hadoop-lzo rev 049362b7cf53ff5f739d6b1532457f2c6cd495e8]
  Total input files to process : 1
  number of splits:4
  Submitting tokens for job: job_1663628638628_0010
  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_1663628638628_0010
  The url to track the job: http://ip-172-31-38-254.us-east-2.compute.internal:20888/proxy/application_1663628638628_0010/
  Running job: job_1663628638628_0010
  Job job_1663628638628_0010 running in uber mode : false
   map 0% reduce 0%
   map 50% reduce 0%
   map 100% reduce 0%
```

```
job output is in hdfs:///user/hadoop/tmp/mrjob/Movies.hadoop.20220920.014542.512809/output
Streaming final output from hdfs:///user/hadoop/tmp/mrjob/Movies.hadoop.20220920.014542.512809/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
"120"	138
"121"	80
"122"	40
"123"	33
"124"	85
"125"	210
"126"	64
"127"	21
"128"	323
"129"	26
"13"	53
"130"	375
"131"	44
"132"	94
"133"	178
"134"	311
"135"	22
"136"	50
"137"	80
"138"	81
"139"	68
"14"	20
"140"	46
"141"	31
"142"	61
"143"	77
"144"	41
"145"	38
"146"	73
"147"	38
"148"	132
"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
Removing HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/Movies.hadoop.20220920.014542.512809...
Removing temp directory /tmp/Movies.hadoop.20220920.014542.512809...
[hadoop@ip-172-31-38-254 ~]$
```