Assignment 2

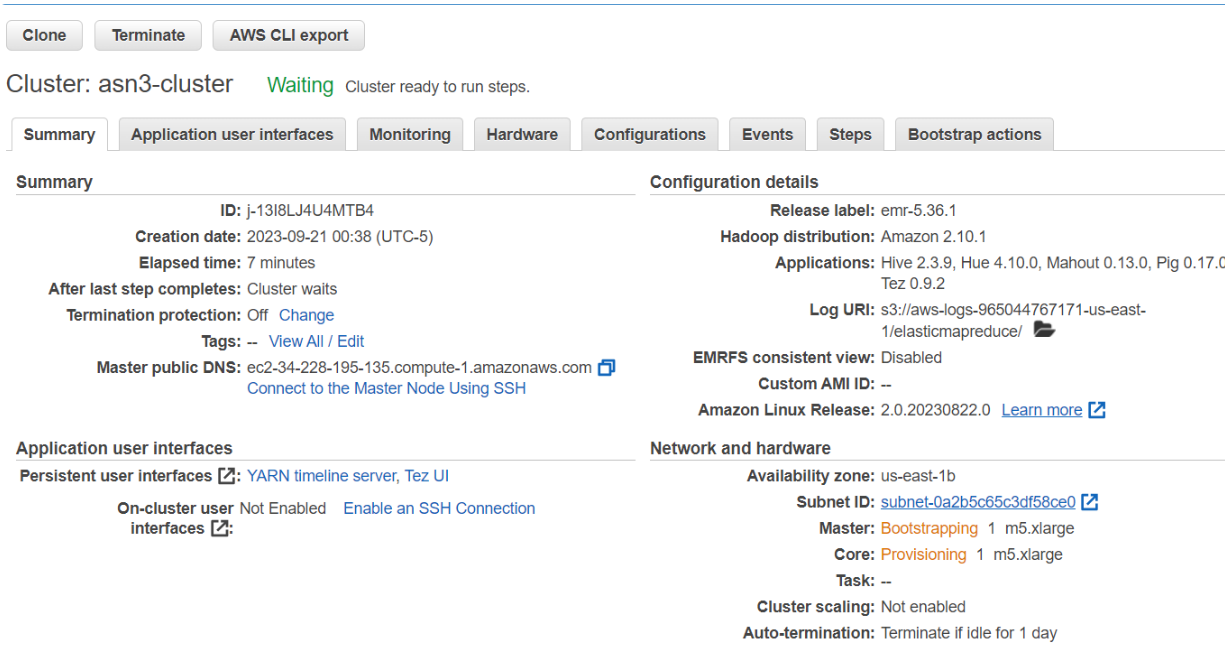
Big Data Technologies – CS554

Prof. Navendu Garg

ID : A20522183

Pem key created, on the EC2 instance

Created EMR – cluster on the console.



Cluster: asn3-cluster Waiting Cluster ready to run steps.

Summary

ID: j-13I8LJ4U4MTB4
Creation date: 2023-09-21 00:38 (UTC-5)
Elapsed time: 7 minutes
After last step completes: Cluster waits
Termination protection: Off Change
Tags: -- View All / Edit
Master public DNS: ec2-34-228-195-135.compute-1.amazonaws.com
Connect to the Master Node Using SSH

Configuration details

Release label: emr-5.36.1
Hadoop distribution: Amazon 2.10.1
Applications: Hive 2.3.9, Hue 4.10.0, Mahout 0.13.0, Pig 0.17.0 Tez 0.9.2
Log URI: s3://aws-logs-965044767171-us-east-1/elasticmapreduce/
EMRFS consistent view: Disabled
Custom AMI ID: --
Amazon Linux Release: 2.0.20230822.0 Learn more

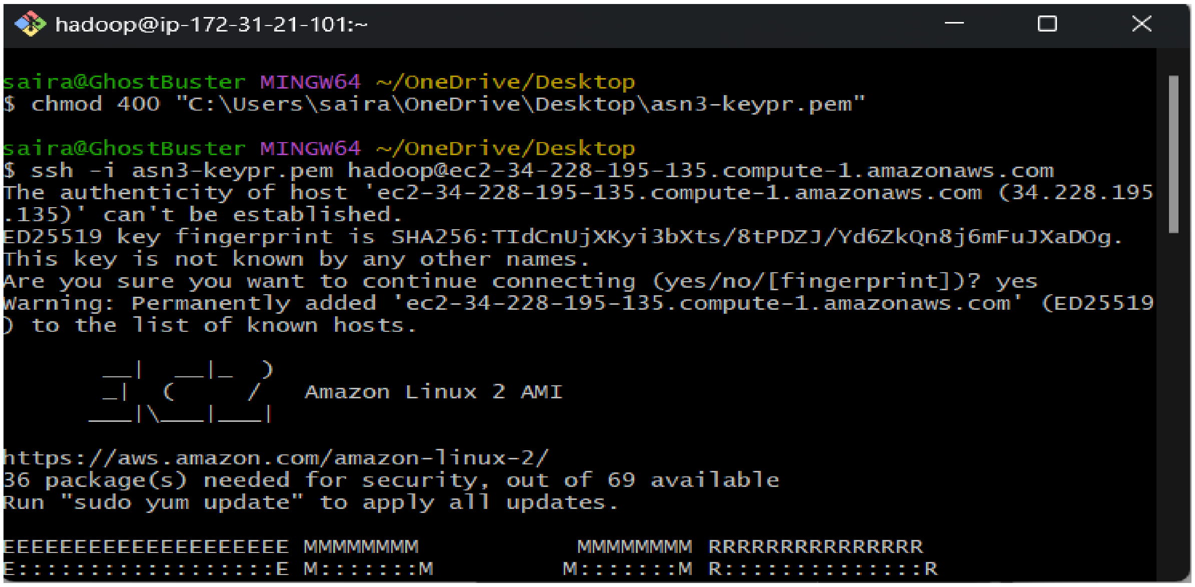
Application user interfaces

Persistent user interfaces: YARN timeline server, Tez UI
On-cluster user interfaces: Not Enabled Enable an SSH Connection

Network and hardware

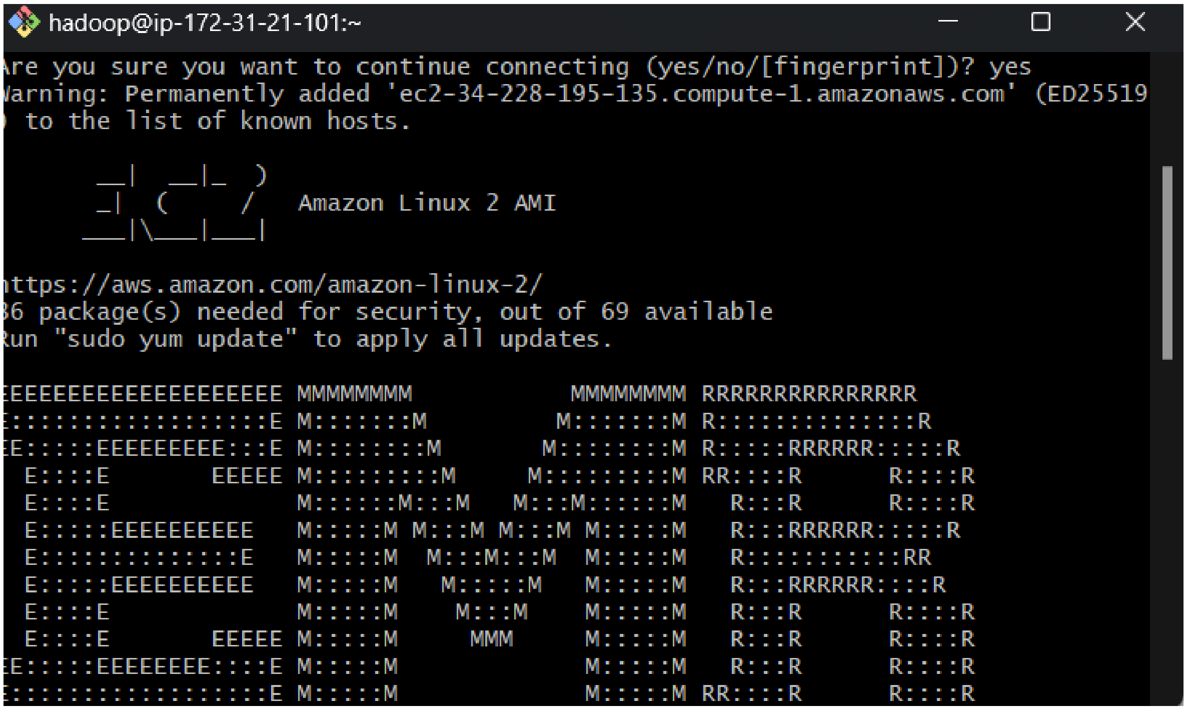
Availability zone: us-east-1b
Subnet ID: subnet-0a2b5c65c3df58ce0
Master: Bootstrapping 1 m5.xlarge
Core: Provisioning 1 m5.xlarge
Task: --
Cluster scaling: Not enabled
Auto-termination: Terminate if idle for 1 day

Modifying permissions of the key-pair directory and connecting to thee Emr – ec2 host



```
saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$ chmod 400 "C:\Users\saira\OneDrive\Desktop\asn3-keypr.pem"

saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$ ssh -i asn3-keypr.pem hadoop@ec2-34-228-195-135.compute-1.amazonaws.com
The authenticity of host 'ec2-34-228-195-135.compute-1.amazonaws.com (34.228.195
.135)' can't be established.
ED25519 key fingerprint is SHA256:TIdCnUjXKyi3bXts/8tPDZJ/Yd6ZkQn8j6mFuJXaDOg.
This key is not known by any other names.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'ec2-34-228-195-135.compute-1.amazonaws.com' (ED25519
) to the list of known hosts.

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

https://aws.amazon.com/amazon-linux-2/
36 package(s) needed for security, out of 69 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMMM           MMMMMMMM RRRRRRRRRRRRRRR
E::::::::::::::::::E M:::::::M         M:::::::M R::::::::::::::R
```

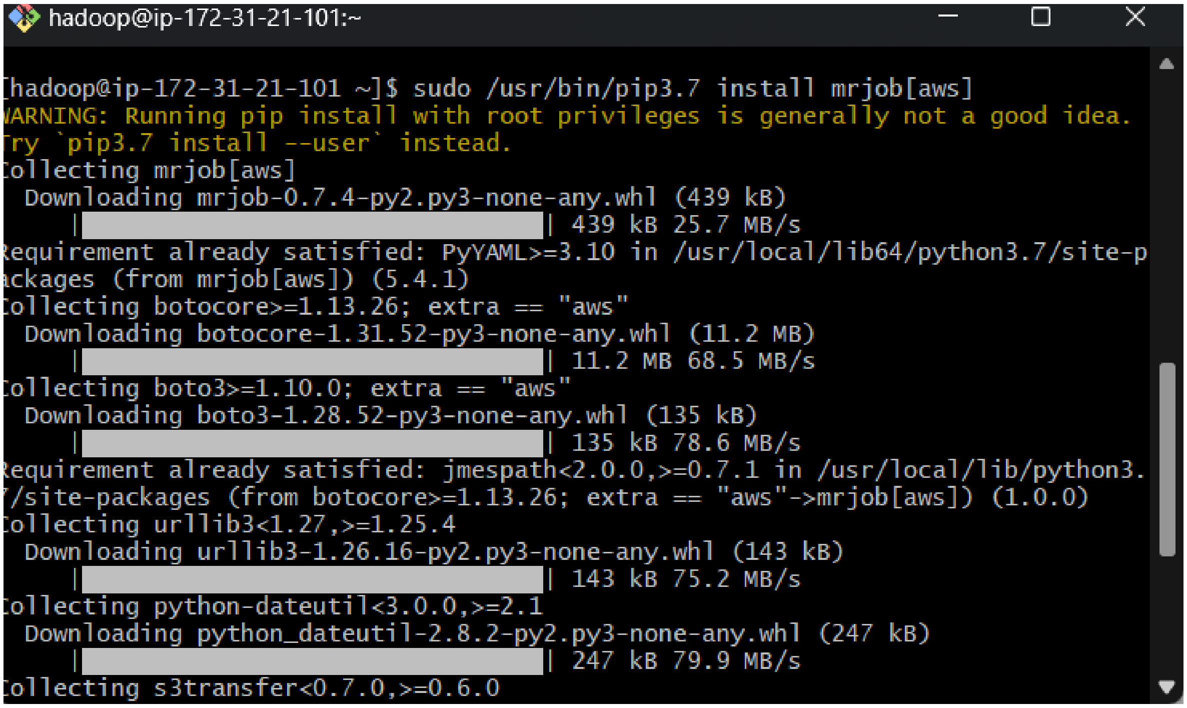
EMR cluster created and connected successfully. Now this we will use to run our hadoop map reduce jobs.



```
re you sure you want to continue connecting (yes/no/[fingerprint])? yes
arning: Permanently added 'ec2-34-228-195-135.compute-1.amazonaws.com' (ED25519
 to the list of known hosts.

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

ttps://aws.amazon.com/amazon-linux-2/
6 package(s) needed for security, out of 69 available
un "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEE MMMMMMMM           MMMMMMMM RRRRRRRRRRRRRRR
:::::::::::::::::::E M:::::::M         M:::::::M R::::::::::::::R
E:::::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
E:::::EEEEEEEE::::E M:::::M             M:::::M   R:::R      R::::R
::::::::::::::::::E M:::::M             M:::::M RR::::R      R::::R
```

Sudo install /usr/bin/pip3.7 install mrjob[aws].

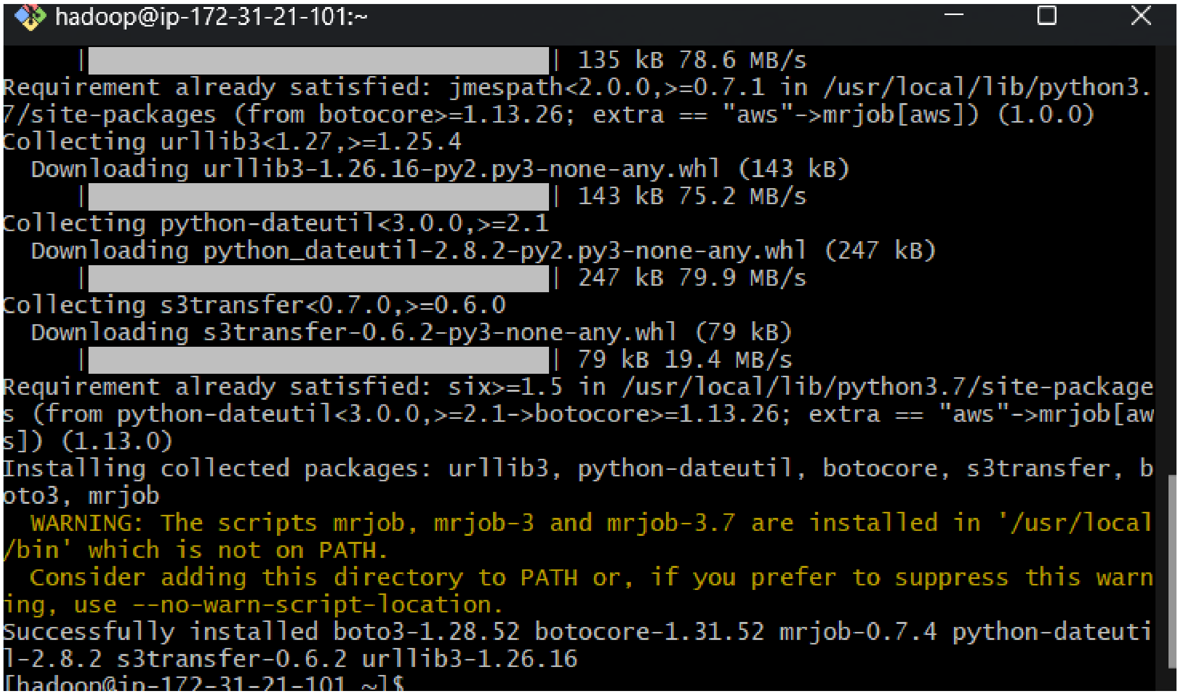
The above command installs mrjob libraries and services for the hadoop map reduce for python files.



hadoop@ip-172-31-21-101:~

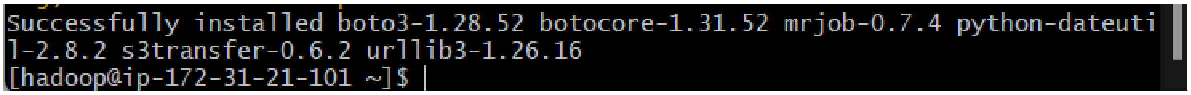
```
[hadoop@ip-172-31-21-101 ~]$ sudo /usr/bin/pip3.7 install mrjob[aws]
WARNING: Running pip install with root privileges is generally not a good idea.
Try `pip3.7 install --user` instead.
Collecting mrjob[aws]
  Downloading mrjob-0.7.4-py2.py3-none-any.whl (439 kB)
     |████████████████████████████████| 439 kB 25.7 MB/s
Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib64/python3.7/site-p
ackages (from mrjob[aws]) (5.4.1)
Collecting botocore>=1.13.26; extra == "aws"
  Downloading botocore-1.31.52-py3-none-any.whl (11.2 MB)
     |████████████████████████████████| 11.2 MB 68.5 MB/s
Collecting boto3>=1.10.0; extra == "aws"
  Downloading boto3-1.28.52-py3-none-any.whl (135 kB)
     |████████████████████████████████| 135 kB 78.6 MB/s
Requirement already satisfied: jmespath<2.0.0,>=0.7.1 in /usr/local/lib/python3.
7/site-packages (from botocore>=1.13.26; extra == "aws"->mrjob[aws]) (1.0.0)
Collecting urllib3<1.27,>=1.25.4
  Downloading urllib3-1.26.16-py2.py3-none-any.whl (143 kB)
     |████████████████████████████████| 143 kB 75.2 MB/s
Collecting python-dateutil<3.0.0,>=2.1
  Downloading python_dateutil-2.8.2-py2.py3-none-any.whl (247 kB)
     |████████████████████████████████| 247 kB 79.9 MB/s
Collecting s3transfer<0.7.0,>=0.6.0
```

Successfully connected to pip python3.7



hadoop@ip-172-31-21-101:~

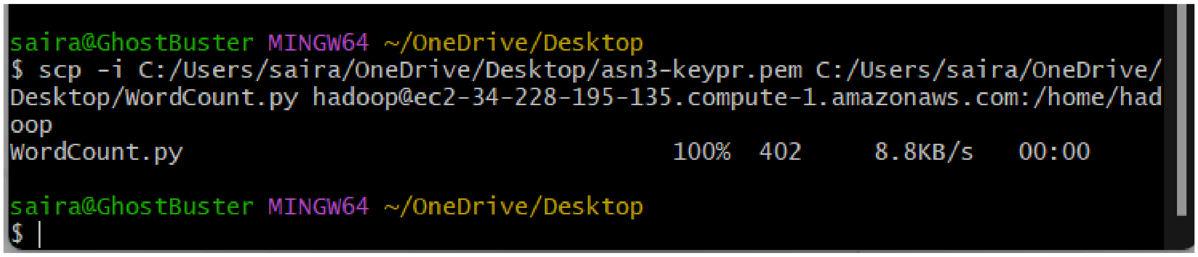
```
     |████████████████████████████████| 135 kB 78.6 MB/s
Requirement already satisfied: jmespath<2.0.0,>=0.7.1 in /usr/local/lib/python3.
7/site-packages (from botocore>=1.13.26; extra == "aws"->mrjob[aws]) (1.0.0)
Collecting urllib3<1.27,>=1.25.4
  Downloading urllib3-1.26.16-py2.py3-none-any.whl (143 kB)
     |████████████████████████████████| 143 kB 75.2 MB/s
Collecting python-dateutil<3.0.0,>=2.1
  Downloading python_dateutil-2.8.2-py2.py3-none-any.whl (247 kB)
     |████████████████████████████████| 247 kB 79.9 MB/s
Collecting s3transfer<0.7.0,>=0.6.0
  Downloading s3transfer-0.6.2-py3-none-any.whl (79 kB)
     |████████████████████████████████| 79 kB 19.4 MB/s
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/site-package
s (from python-dateutil<3.0.0,>=2.1->botocore>=1.13.26; extra == "aws"->mrjob[aw
s]) (1.13.0)
Installing collected packages: urllib3, python-dateutil, botocore, s3transfer, b
oto3, 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 warn
ing, use --no-warn-script-location.
Successfully installed boto3-1.28.52 botocore-1.31.52 mrjob-0.7.4 python-dateuti
l-2.8.2 s3transfer-0.6.2 urllib3-1.26.16
[hadoop@ip-172-31-21-101 ~]$
```



```
Successfully installed boto3-1.28.52 botocore-1.31.52 mrjob-0.7.4 python-dateuti
l-2.8.2 s3transfer-0.6.2 urllib3-1.26.16
[hadoop@ip-172-31-21-101 ~]$ |
```

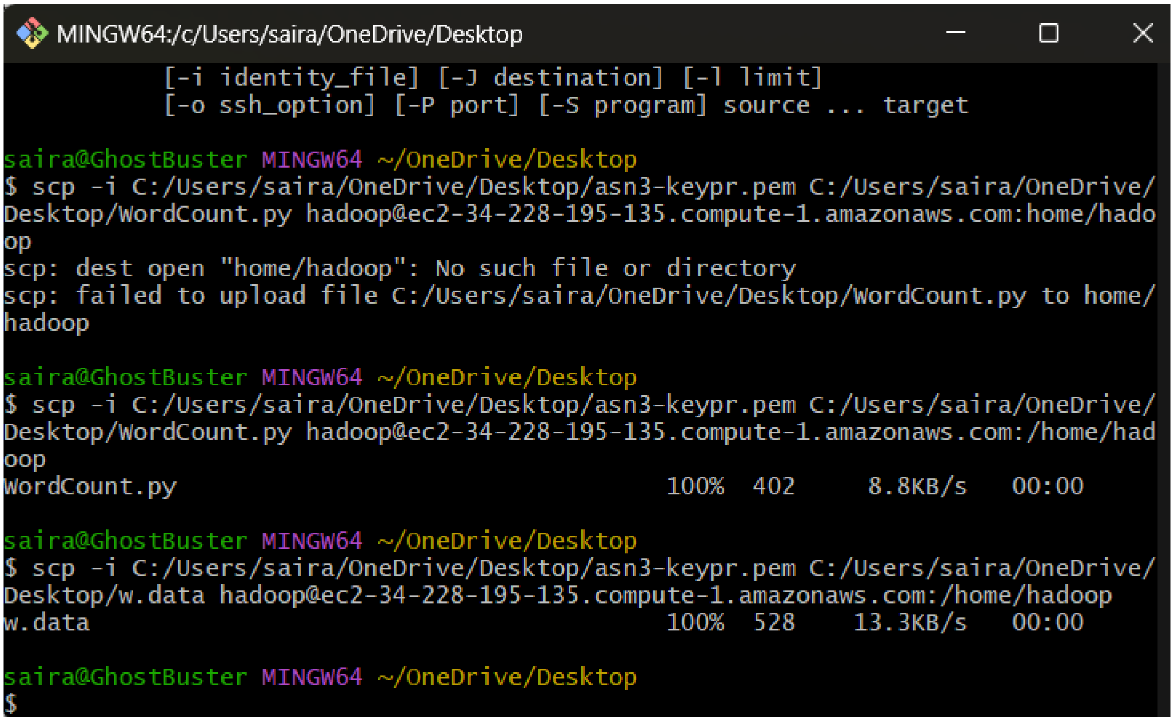
Step 4

Word count py , is securely copied using the command **scp -i /path/to/.pem /path/to/wordcount.py**



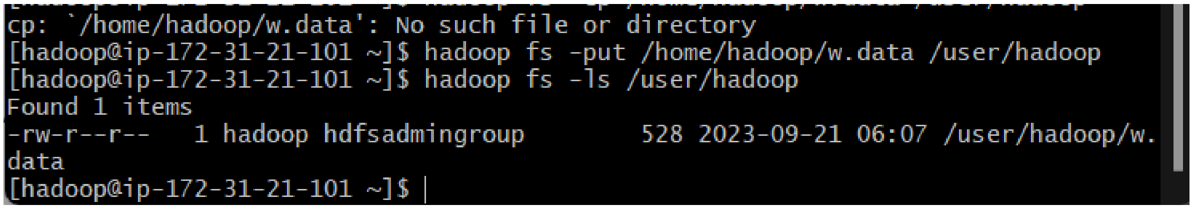
```
saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$ scp -i C:/Users/saira/OneDrive/Desktop/asn3-keypr.pem C:/Users/saira/OneDrive/
Desktop/WordCount.py hadoop@ec2-34-228-195-135.compute-1.amazonaws.com:/home/had
oop
WordCount.py                                  100%  402     8.8KB/s   00:00

saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$ |
```

scp -i /path/to/.pem /path/to/w.data



```
           [-i identity_file] [-J destination] [-l limit]
           [-o ssh_option] [-P port] [-S program] source ... target

saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$ scp -i C:/Users/saira/OneDrive/Desktop/asn3-keypr.pem C:/Users/saira/OneDrive/
Desktop/WordCount.py hadoop@ec2-34-228-195-135.compute-1.amazonaws.com:home/hado
op
scp: dest open "home/hadoop": No such file or directory
scp: failed to upload file C:/Users/saira/OneDrive/Desktop/WordCount.py to home/
hadoop

saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$ scp -i C:/Users/saira/OneDrive/Desktop/asn3-keypr.pem C:/Users/saira/OneDrive/
Desktop/WordCount.py hadoop@ec2-34-228-195-135.compute-1.amazonaws.com:/home/had
oop
WordCount.py                                  100%  402     8.8KB/s   00:00

saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$ scp -i C:/Users/saira/OneDrive/Desktop/asn3-keypr.pem C:/Users/saira/OneDrive/
Desktop/w.data hadoop@ec2-34-228-195-135.compute-1.amazonaws.com:/home/hadoop
w.data                                        100%  528    13.3KB/s   00:00

saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$
```

Checking if the w.data file is in the cluster using hadoop fs -ls/user/hadoop

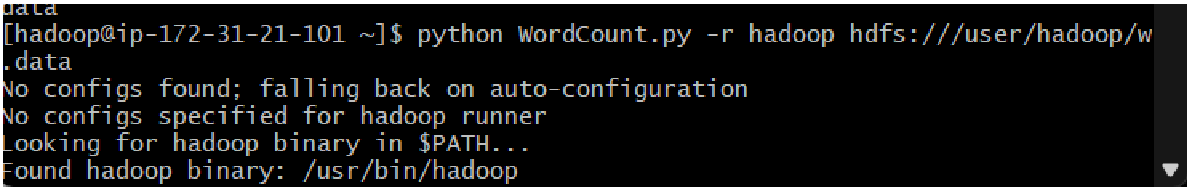


```
cp: `/home/hadoop/w.data': No such file or directory
[hadoop@ip-172-31-21-101 ~]$ hadoop fs -put /home/hadoop/w.data /user/hadoop
[hadoop@ip-172-31-21-101 ~]$ hadoop fs -ls /user/hadoop
Found 1 items
-rw-r--r--   1 hadoop hdfsadmingroup        528 2023-09-21 06:07 /user/hadoop/w.
data
[hadoop@ip-172-31-21-101 ~]$ |
```

Successfully file is uploaded onto the cluster.

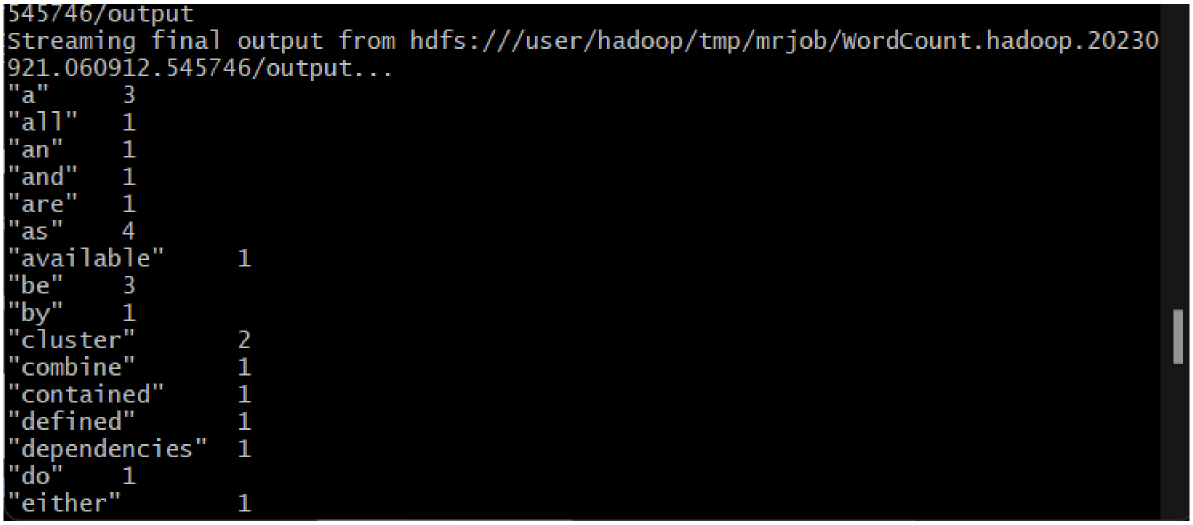
Now we run the python wordcount.py -r hadoop hdfs:////user/hadoop/w.data

This will take w.data as input data into the wordcount py and the MapReduce job will begin to run.

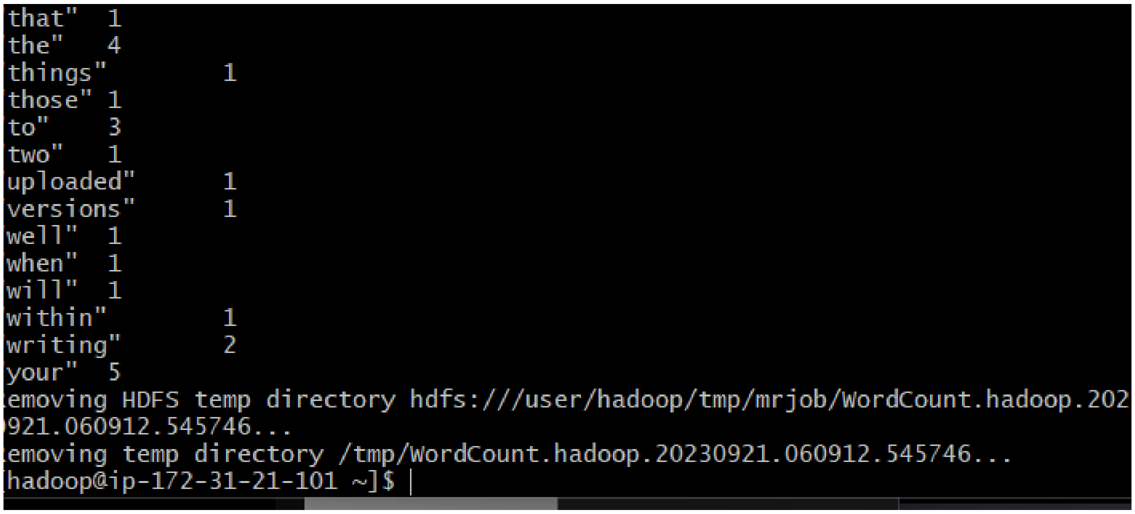


```
data
[hadoop@ip-172-31-21-101 ~]$ python WordCount.py -r hadoop hdfs:///user/hadoop/w
.data
No configs found; falling back on auto-configuration
No configs specified for hadoop runner
Looking for hadoop binary in $PATH...
Found hadoop binary: /usr/bin/hadoop
```

As we can see the output has words – word frequency

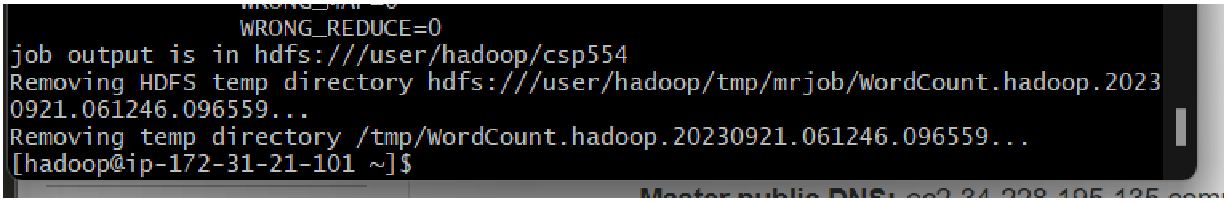


```
545746/output
Streaming final output from hdfs:///user/hadoop/tmp/mrjob/WordCount.hadoop.20230
921.060912.545746/output...
"a"     3
"all"   1
"an"    1
"and"   1
"are"   1
"as"    4
"available"     1
"be"    3
"by"    1
"cluster"       2
"combine"       1
"contained"     1
"defined"       1
"dependencies"  1
"do"    1
"either"        1
```



```
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
emoving HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/WordCount.hadoop.202
921.060912.545746...
emoving temp directory /tmp/WordCount.hadoop.20230921.060912.545746...
hadoop@ip-172-31-21-101 ~]$ |
```

Now we move the output into new directory



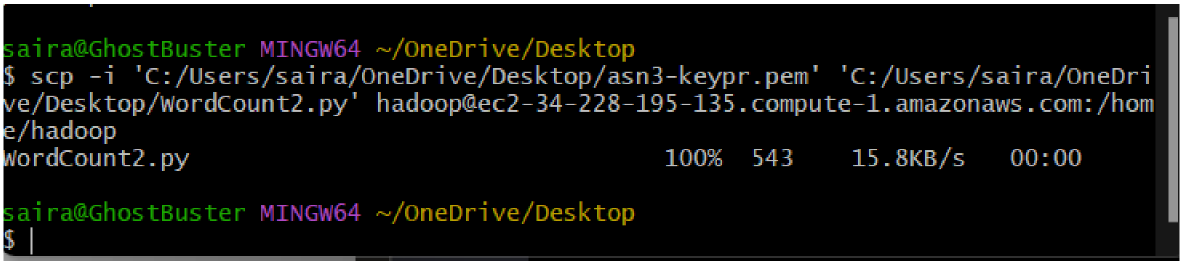
```
                WRONG_MAP=0
                WRONG_REDUCE=0
job output is in hdfs:///user/hadoop/csp554
Removing HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/WordCount.hadoop.2023
0921.061246.096559...
Removing temp directory /tmp/WordCount.hadoop.20230921.061246.096559...
[hadoop@ip-172-31-21-101 ~]$
```

Master public DNS: ec2-34-228-195-135.com

Step 5.

Word count2 py

Scp -i /path/to/pem/file /wordcount2.py

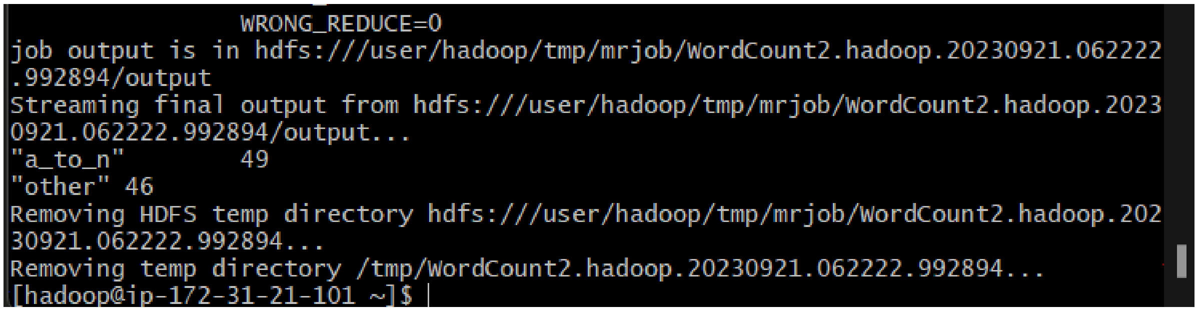


```
saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$ scp -i 'C:/Users/saira/OneDrive/Desktop/asn3-keypr.pem' 'C:/Users/saira/OneDri
ve/Desktop/WordCount2.py' hadoop@ec2-34-228-195-135.compute-1.amazonaws.com:/hom
e/hadoop
WordCount2.py                                 100%  543     15.8KB/s   00:00

saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$ |
```

6. WordCount2.py output here **a_to_n 49 ,**

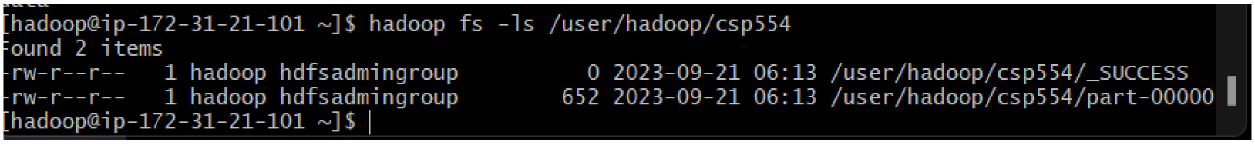
Other 46

Output in new directory csp554



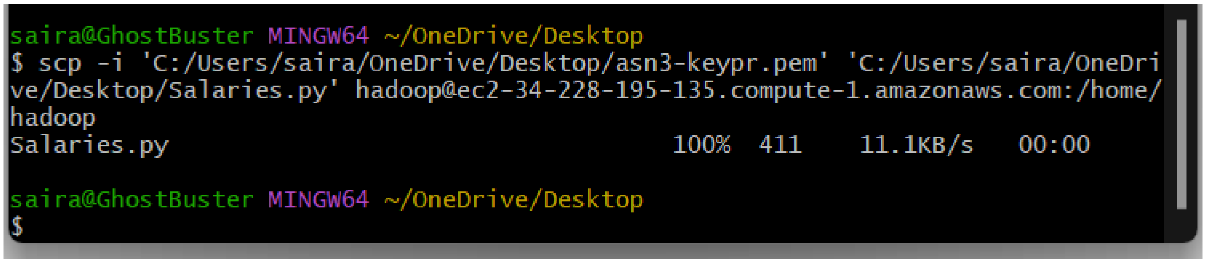
```
                WRONG_REDUCE=0
job output is in hdfs:///user/hadoop/tmp/mrjob/WordCount2.hadoop.20230921.062222
.992894/output
Streaming final output from hdfs:///user/hadoop/tmp/mrjob/WordCount2.hadoop.2023
0921.062222.992894/output...
"a_to_n"        49
"other" 46
Removing HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/WordCount2.hadoop.202
30921.062222.992894...
Removing temp directory /tmp/WordCount2.hadoop.20230921.062222.992894...
[hadoop@ip-172-31-21-101 ~]$ |
```

Both programs are stored in /user/hadoop/csp554



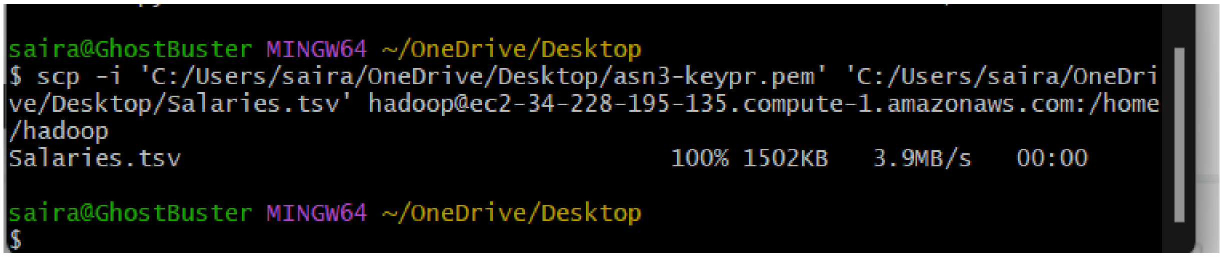
```
[hadoop@ip-172-31-21-101 ~]$ hadoop fs -ls /user/hadoop/csp554
Found 2 items
-rw-r--r--   1 hadoop hdfsadmingroup          0 2023-09-21 06:13 /user/hadoop/csp554/_SUCCESS
-rw-r--r--   1 hadoop hdfsadmingroup        652 2023-09-21 06:13 /user/hadoop/csp554/part-00000
[hadoop@ip-172-31-21-101 ~]$ |
```

Salaries.py



```
saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$ scp -i 'C:/Users/saira/OneDrive/Desktop/asn3-keypr.pem' 'C:/Users/saira/OneDri
ve/Desktop/Salaries.py' hadoop@ec2-34-228-195-135.compute-1.amazonaws.com:/home/
hadoop
Salaries.py                                   100%  411     11.1KB/s   00:00

saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$
```

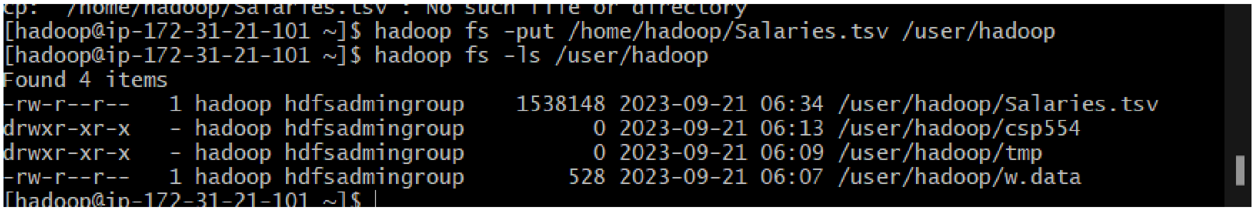
Upload salaries.tsv



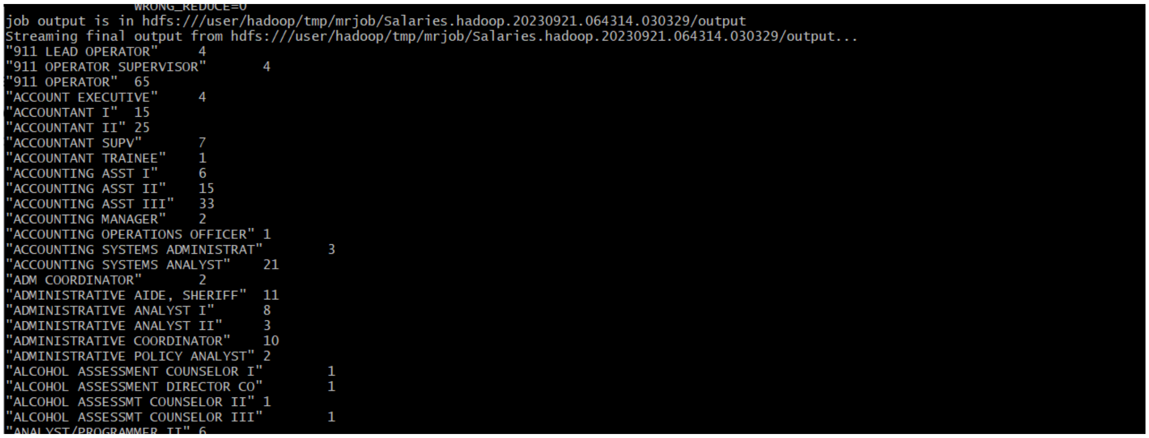
```
saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$ scp -i 'C:/Users/saira/OneDrive/Desktop/asn3-keypr.pem' 'C:/Users/saira/OneDrive/Desktop/Salaries.tsv' hadoop@ec2-34-228-195-135.compute-1.amazonaws.com:/home/hadoop
Salaries.tsv                                  100% 1502KB   3.9MB/s   00:00

saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$
```

Step 7.

Moving tsv into home hadoop

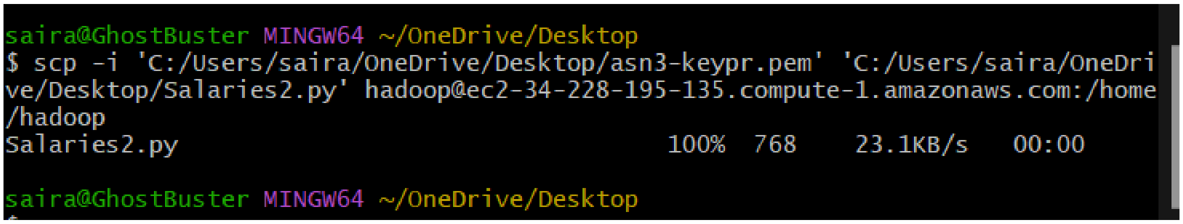


```
cp: '/home/hadoop/Salaries.tsv': No such file or directory
[hadoop@ip-172-31-21-101 ~]$ hadoop fs -put /home/hadoop/Salaries.tsv /user/hadoop
[hadoop@ip-172-31-21-101 ~]$ hadoop fs -ls /user/hadoop
Found 4 items
-rw-r--r--   1 hadoop hdfsadmingroup    1538148 2023-09-21 06:34 /user/hadoop/Salaries.tsv
drwxr-xr-x   - hadoop hdfsadmingroup          0 2023-09-21 06:13 /user/hadoop/csp554
drwxr-xr-x   - hadoop hdfsadmingroup          0 2023-09-21 06:09 /user/hadoop/tmp
-rw-r--r--   1 hadoop hdfsadmingroup        528 2023-09-21 06:07 /user/hadoop/w.data
[hadoop@ip-172-31-21-101 ~]$
```



```
                WRONG_REDUCE=0
job output is in hdfs:///user/hadoop/tmp/mrjob/Salaries.hadoop.20230921.064314.030329/output
Streaming final output from hdfs:///user/hadoop/tmp/mrjob/Salaries.hadoop.20230921.064314.030329/output...
"911 LEAD OPERATOR"     4
"911 OPERATOR SUPERVISOR"       4
"911 OPERATOR"  65
"ACCOUNT EXECUTIVE"     4
"ACCOUNTANT I"  15
"ACCOUNTANT II" 25
"ACCOUNTANT SUPV"       7
"ACCOUNTANT TRAINEE"    1
"ACCOUNTING ASST I"     6
"ACCOUNTING ASST II"    15
"ACCOUNTING ASST III"   33
"ACCOUNTING MANAGER"    2
"ACCOUNTING OPERATIONS OFFICER" 1
"ACCOUNTING SYSTEMS ADMINISTRAT"        3
"ACCOUNTING SYSTEMS ANALYST"    21
"ADM COORDINATOR"       2
"ADMINISTRATIVE AIDE, SHERIFF"  11
"ADMINISTRATIVE ANALYST I"      8
"ADMINISTRATIVE ANALYST II"     3
"ADMINISTRATIVE COORDINATOR"    10
"ADMINISTRATIVE POLICY ANALYST" 2
"ALCOHOL ASSESSMENT COUNSELOR I"        1
"ALCOHOL ASSESSMENT DIRECTOR CO"        1
"ALCOHOL ASSESSMT COUNSELOR II" 1
"ALCOHOL ASSESSMT COUNSELOR III"        1
"ANALYST/PROGRAMMER II" 6
```

9.) Salaries2.py

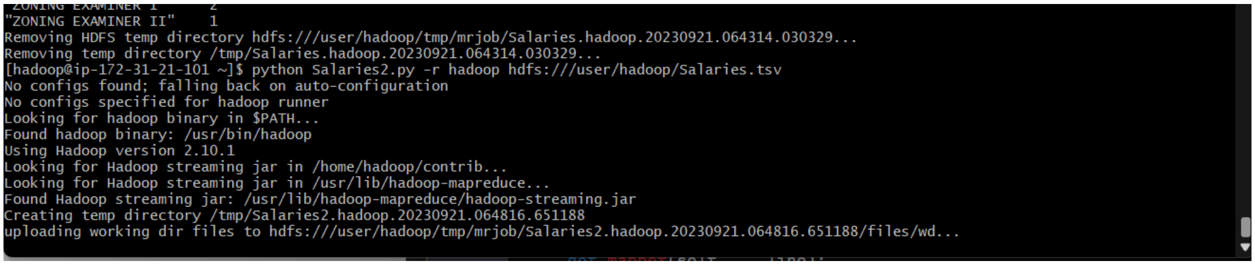


```
saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$ scp -i 'C:/Users/saira/OneDrive/Desktop/asn3-keypr.pem' 'C:/Users/saira/OneDrive/Desktop/Salaries2.py' hadoop@ec2-34-228-195-135.compute-1.amazonaws.com:/home/hadoop
Salaries2.py                                  100%  768     23.1KB/s   00:00

saira@GhostBuster MINGW64 ~/OneDrive/Desktop
```

Salaries2 py with salaries.tsv

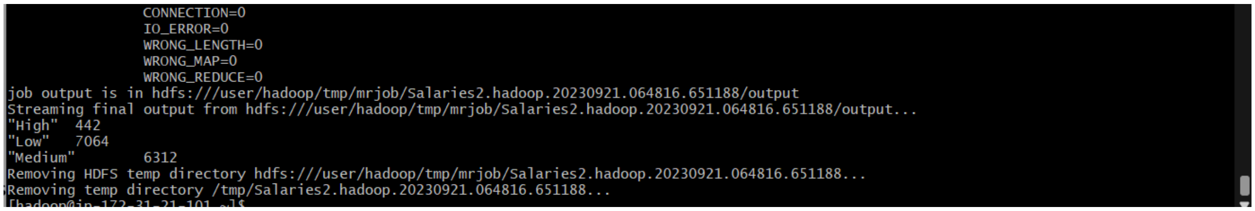
Python salaries2.py -r hadoop hdfs:///user/hadoop/salaries.tsv

execute the command above



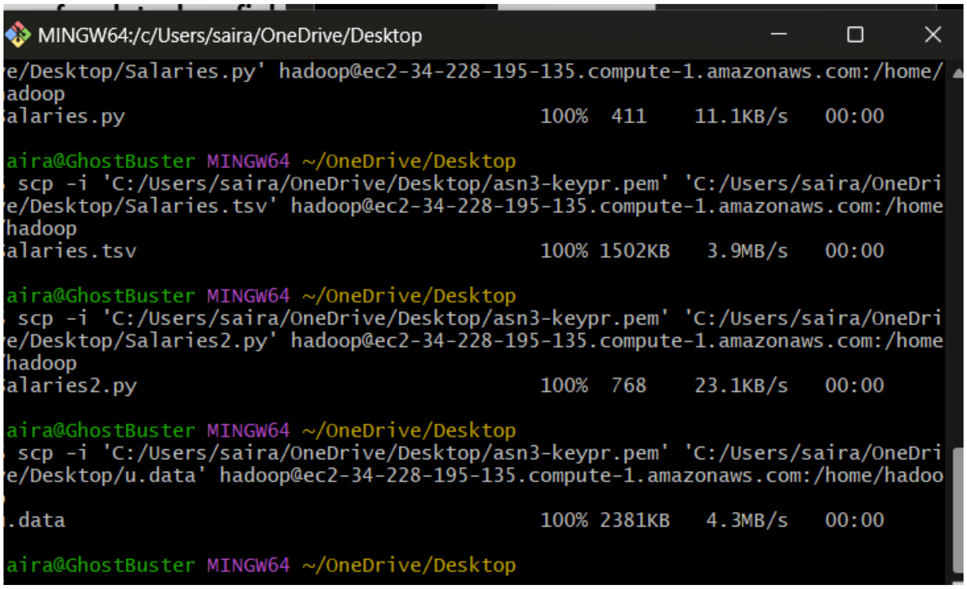
```
"ZONING EXAMINER I"      2
"ZONING EXAMINER II"    1
Removing HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/Salaries.hadoop.20230921.064314.030329...
Removing temp directory /tmp/Salaries.hadoop.20230921.064314.030329...
[hadoop@ip-172-31-21-101 ~]$ python Salaries2.py -r hadoop hdfs:///user/hadoop/Salaries.tsv
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 2.10.1
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.20230921.064816.651188
uploading working dir files to hdfs:///user/hadoop/tmp/mrjob/Salaries2.hadoop.20230921.064816.651188/files/wd...
```

11). Count of employees output is on the terminal screen, that means our job is successful



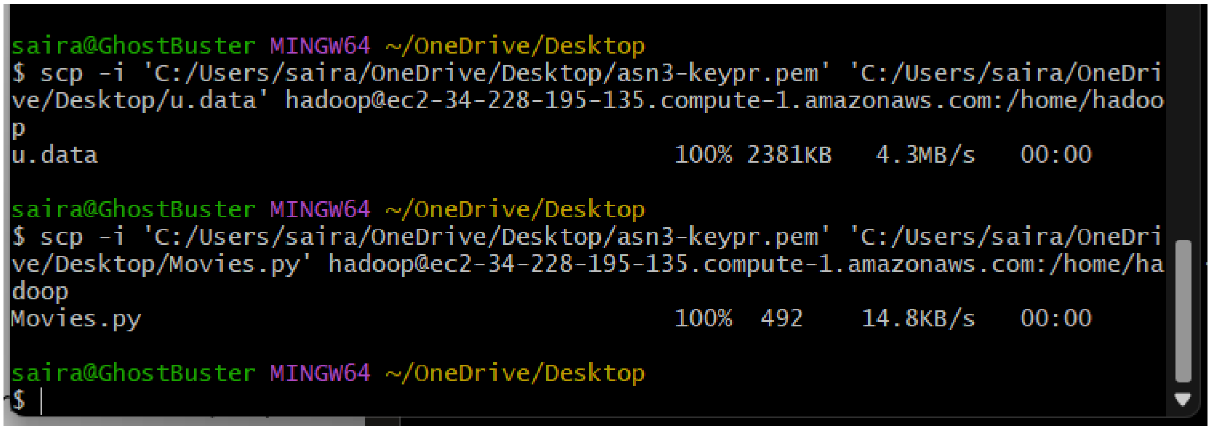
```
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
job output is in hdfs:///user/hadoop/tmp/mrjob/Salaries2.hadoop.20230921.064816.651188/output
Streaming final output from hdfs:///user/hadoop/tmp/mrjob/Salaries2.hadoop.20230921.064816.651188/output...
"High"  442
"Low"   7064
"Medium"        6312
Removing HDFS temp directory hdfs:///user/hadoop/tmp/mrjob/Salaries2.hadoop.20230921.064816.651188...
Removing temp directory /tmp/Salaries2.hadoop.20230921.064816.651188...
[hadoop@ip-172-31-21-101 ~]$
```

12). Scp u.data



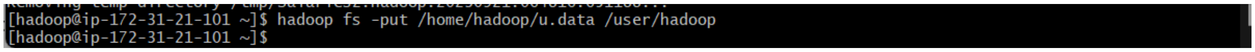
```
e/Desktop/Salaries.py' hadoop@ec2-34-228-195-135.compute-1.amazonaws.com:/home/
adoop
alaries.py                                    100%  411     11.1KB/s   00:00

aira@GhostBuster MINGW64 ~/OneDrive/Desktop
 scp -i 'C:/Users/saira/OneDrive/Desktop/asn3-keypr.pem' 'C:/Users/saira/OneDri
e/Desktop/Salaries.tsv' hadoop@ec2-34-228-195-135.compute-1.amazonaws.com:/home
hadoop
alaries.tsv                                   100% 1502KB   3.9MB/s   00:00

aira@GhostBuster MINGW64 ~/OneDrive/Desktop
 scp -i 'C:/Users/saira/OneDrive/Desktop/asn3-keypr.pem' 'C:/Users/saira/OneDri
e/Desktop/Salaries2.py' hadoop@ec2-34-228-195-135.compute-1.amazonaws.com:/home
hadoop
alaries2.py                                   100%  768     23.1KB/s   00:00

aira@GhostBuster MINGW64 ~/OneDrive/Desktop
 scp -i 'C:/Users/saira/OneDrive/Desktop/asn3-keypr.pem' 'C:/Users/saira/OneDri
e/Desktop/u.data' hadoop@ec2-34-228-195-135.compute-1.amazonaws.com:/home/hadoo

.data                                         100% 2381KB   4.3MB/s   00:00

aira@GhostBuster MINGW64 ~/OneDrive/Desktop
```

13).Scp movies.py



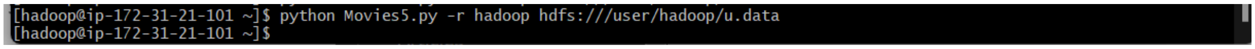
```
saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$ scp -i 'C:/Users/saira/OneDrive/Desktop/asn3-keypr.pem' 'C:/Users/saira/OneDri
ve/Desktop/u.data' hadoop@ec2-34-228-195-135.compute-1.amazonaws.com:/home/hadoo
p
u.data                                        100% 2381KB   4.3MB/s   00:00

saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$ scp -i 'C:/Users/saira/OneDrive/Desktop/asn3-keypr.pem' 'C:/Users/saira/OneDri
ve/Desktop/Movies.py' hadoop@ec2-34-228-195-135.compute-1.amazonaws.com:/home/ha
doop
Movies.py                                     100%  492    14.8KB/s   00:00

saira@GhostBuster MINGW64 ~/OneDrive/Desktop
$ |
```

Hadoop fs -put /home/hadoop/u.data /user/hadoop



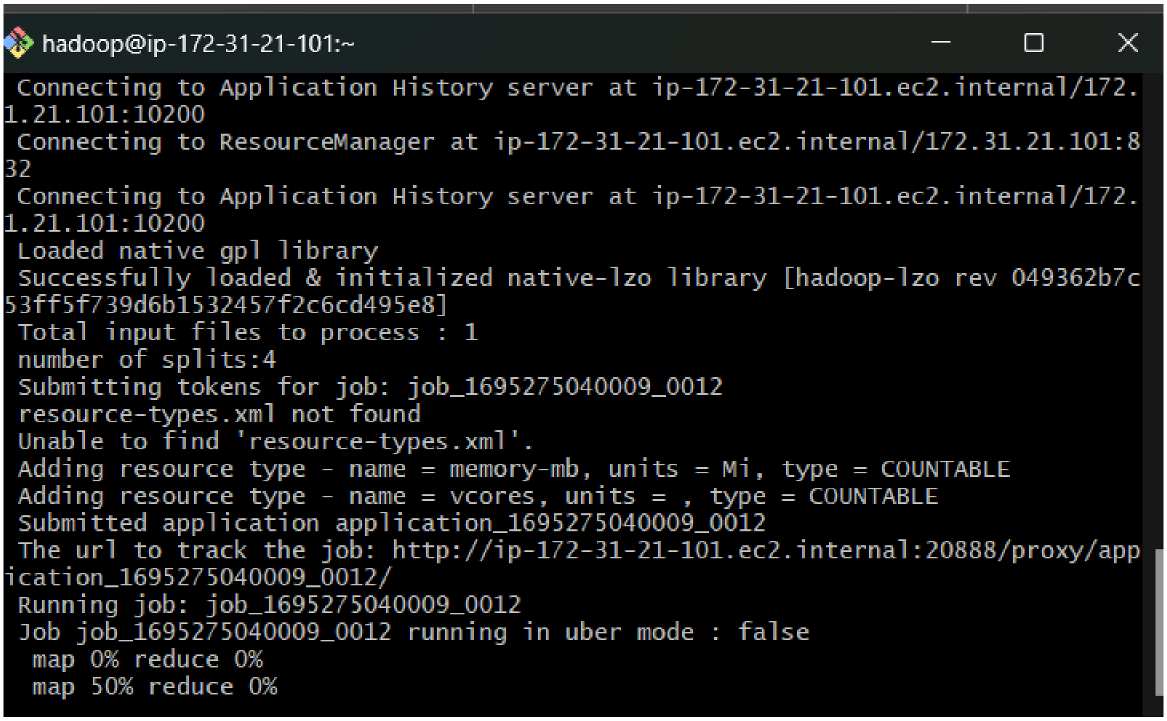
```
[hadoop@ip-172-31-21-101 ~]$ hadoop fs -put /home/hadoop/u.data /user/hadoop
[hadoop@ip-172-31-21-101 ~]$
```

Now execute python movies.py hadoop hdfs:///user/hadoop/u.data



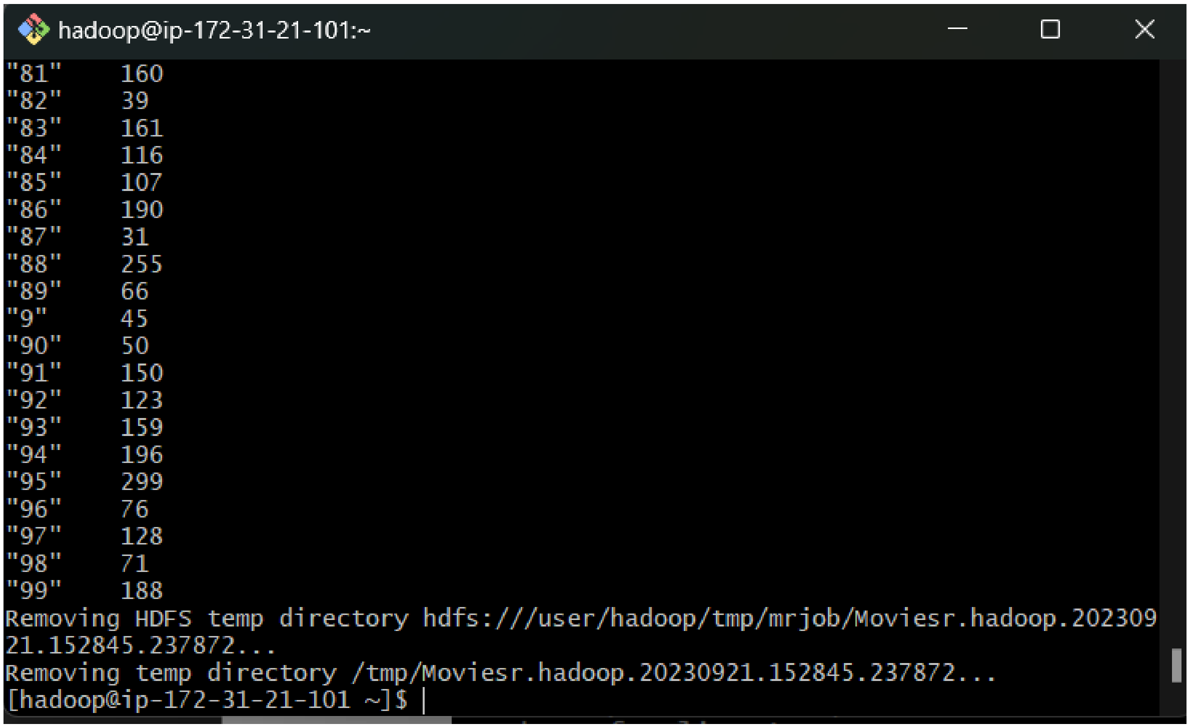
```
[hadoop@ip-172-31-21-101 ~]$ python Movies5.py -r hadoop hdfs:///user/hadoop/u.data
[hadoop@ip-172-31-21-101 ~]$
```

python Movies.py -r hadoop hdfs:///user/hadoop/u.data



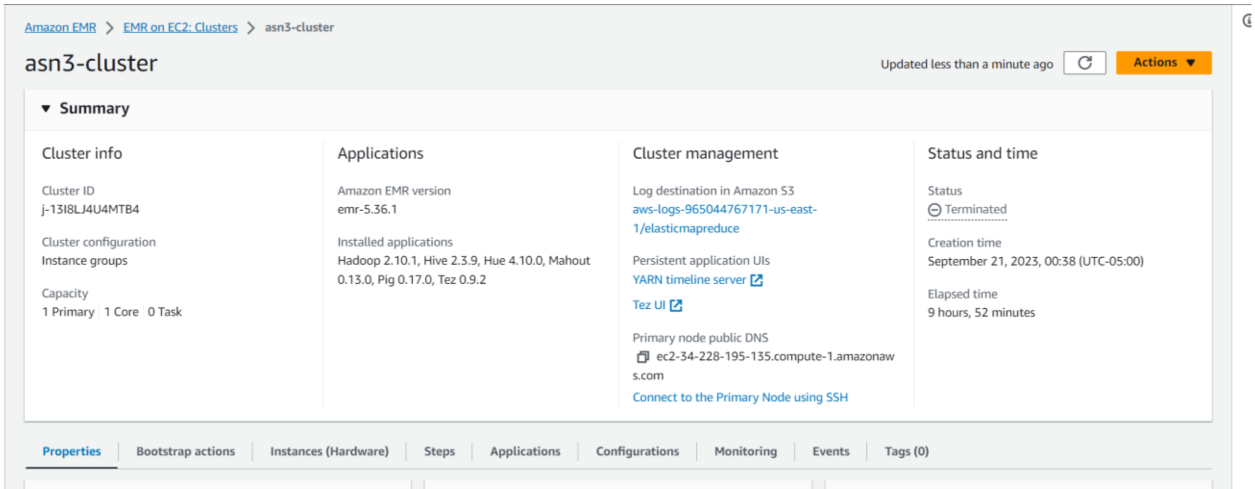
```
 Connecting to Application History server at ip-172-31-21-101.ec2.internal/172.
1.21.101:10200
 Connecting to ResourceManager at ip-172-31-21-101.ec2.internal/172.31.21.101:8
32
 Connecting to Application History server at ip-172-31-21-101.ec2.internal/172.
1.21.101:10200
 Loaded native gpl library
 Successfully loaded & initialized native-lzo library [hadoop-lzo rev 049362b7c
53ff5f739d6b1532457f2c6cd495e8]
 Total input files to process : 1
 number of splits:4
 Submitting tokens for job: job_1695275040009_0012
 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_1695275040009_0012
 The url to track the job: http://ip-172-31-21-101.ec2.internal:20888/proxy/app
ication_1695275040009_0012/
 Running job: job_1695275040009_0012
 Job job_1695275040009_0012 running in uber mode : false
  map 0% reduce 0%
  map 50% reduce 0%
```

Output of the Movies job.



```
"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/Moviesr.hadoop.202309
21.152845.237872...
Removing temp directory /tmp/Moviesr.hadoop.20230921.152845.237872...
[hadoop@ip-172-31-21-101 ~]$ |
```

Finally Cluster terminated.



Amazon EMR > EMR on EC2: Clusters > asn3-cluster
Updated less than a minute ago
Actions ▼
▼ Summary

asn3-cluster

Cluster info

Cluster ID
j-13I8LJ4U4MTB4

Cluster configuration
Instance groups

Capacity
1 Primary | 1 Core | 0 Task

Applications

Amazon EMR version
emr-5.36.1

Installed applications
Hadoop 2.10.1, Hive 2.3.9, Hue 4.10.0, Mahout 0.13.0, Pig 0.17.0, Tez 0.9.2

Cluster management

Log destination in Amazon S3
aws-logs-965044767171-us-east-1/elasticmapreduce

Persistent application UIs
YARN timeline server
Tez UI

Primary node public DNS
ec2-34-228-195-135.compute-1.amazonaws.com

Connect to the Primary Node using SSH

Status and time

Status
⊖ Terminated

Creation time
September 21, 2023, 00:38 (UTC-05:00)

Elapsed time
9 hours, 52 minutes

And resources are removed.

--END--