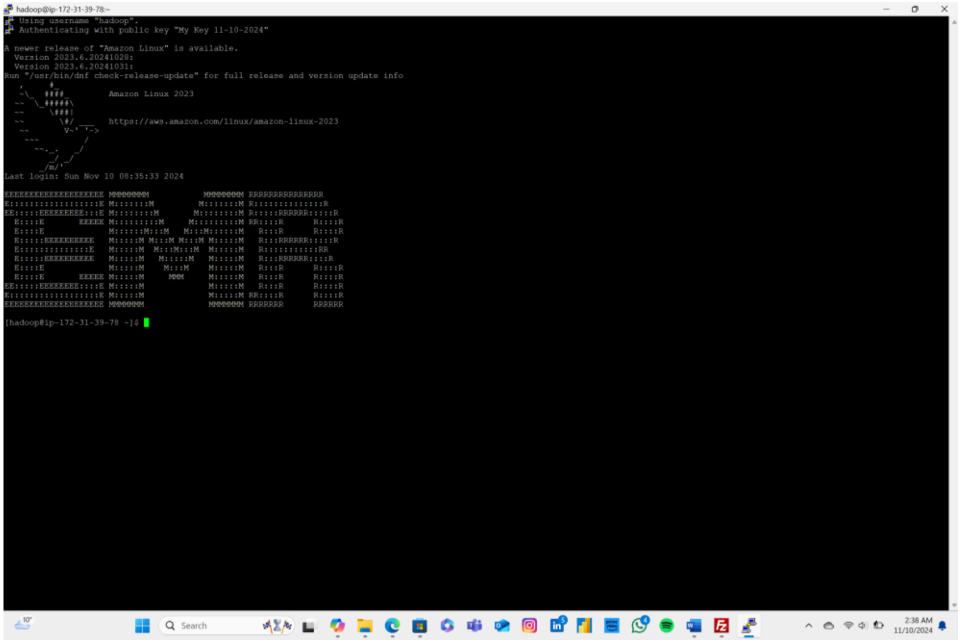
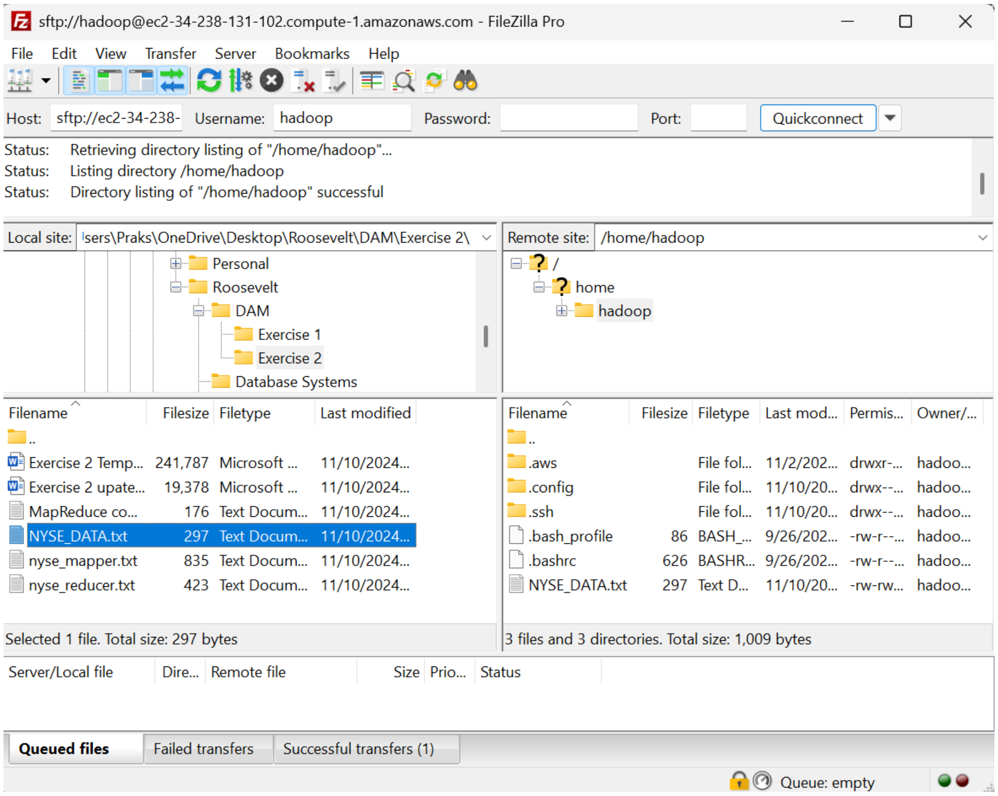
Hadoop MapReduce stock price analysis

1.Task: Clone an EMR and connect to it



```
Using username "hadoop".
Authenticating with public key "My Key 11-10-2024"

A newer release of "Amazon Linux" is available.
  Version 2023.6.20241028:
  Version 2023.6.20241031:
Run "/usr/bin/dnf check-release-update" for full release and version update info
   ,     #_
   ~\_  ####_        Amazon Linux 2023
  ~~  \_#####\
  ~~     \###|
  ~~       \#/ ___   https://aws.amazon.com/linux/amazon-linux-2023
   ~~       V~' '->
    ~~~         /
      ~~._.   _/
         _/ _/
       _/m/'
Last login: Sun Nov 10 08:35:33 2024

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 MMMMMMM             MMMMMMM RRRRRRR      RRRRRR

[hadoop@ip-172-31-39-78 ~]$
```

2. Login and configure WINSCP or FileZilla Pro.



sftp://hadoop@ec2-34-238-131-102.compute-1.amazonaws.com - FileZilla Pro

Host: sftp://ec2-34-238- Username: hadoop Password: Port: Quickconnect

```
Status:    Retrieving directory listing of "/home/hadoop"...
Status:    Listing directory /home/hadoop
Status:    Directory listing of "/home/hadoop" successful
```

Local site: \Users\Praks\OneDrive\Desktop\Roosevelt\DAM\Exercise 2\

- Personal
- Roosevelt
  - DAM
    - Exercise 1
    - Exercise 2
  - Database Systems

Remote site: /home/hadoop

- /
  - home
    - hadoop

Filename	Filesize	Filetype	Last modified
..			
Exercise 2 Temp...	241,787	Microsoft ...	11/10/2024...
Exercise 2 upate...	19,378	Microsoft ...	11/10/2024...
MapReduce co...	176	Text Docum...	11/10/2024...
NYSE_DATA.txt	297	Text Docum...	11/10/2024...
nyse_mapper.txt	835	Text Docum...	11/10/2024...
nyse_reducer.txt	423	Text Docum...	11/10/2024...

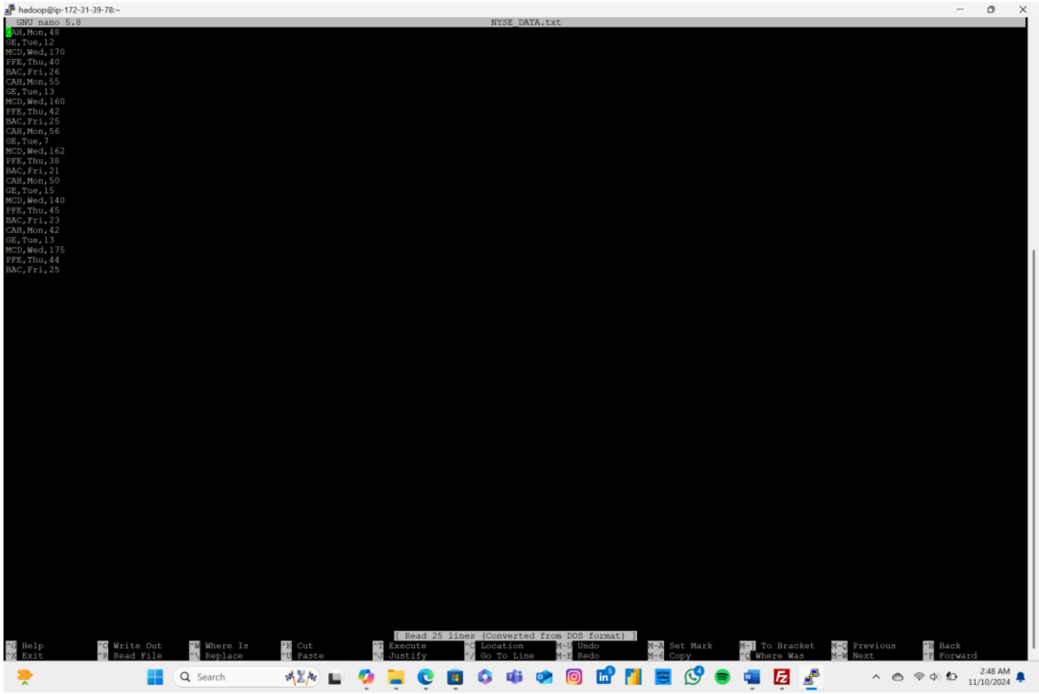
Selected 1 file. Total size: 297 bytes

Filename	Filesize	Filetype	Last mod...	Permis...	Owner/...
..					
.aws		File fol...	11/2/202...	drwxr-...	hadoo...
.config		File fol...	11/10/20...	drwx--...	hadoo...
.ssh		File fol...	11/10/20...	drwx--...	hadoo...
.bash_profile	86	BASH_...	9/26/202...	-rw-r--...	hadoo...
.bashrc	626	BASHR...	9/26/202...	-rw-r--...	hadoo...
NYSE_DATA.txt	297	Text D...	11/10/20...	-rw-rw...	hadoo...

3 files and 3 directories. Total size: 1,009 bytes

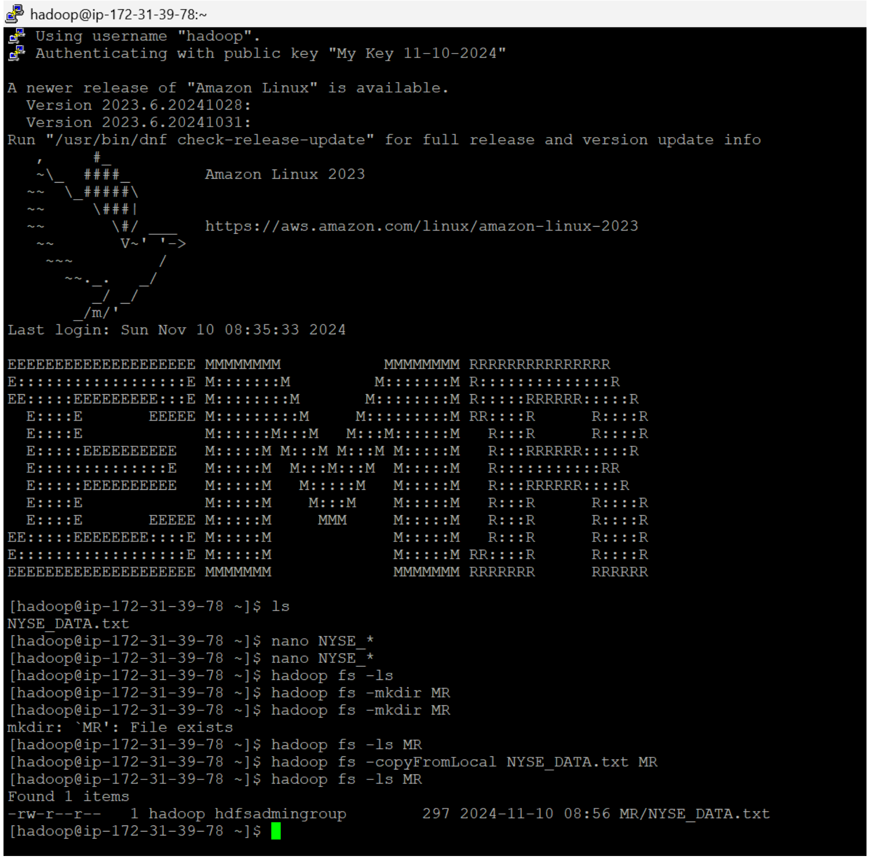
Server/Local file	Dire...	Remote file	Size	Prio...	Status

3. Upload the data file NYSE_DATA.txt to the primary node in EMR



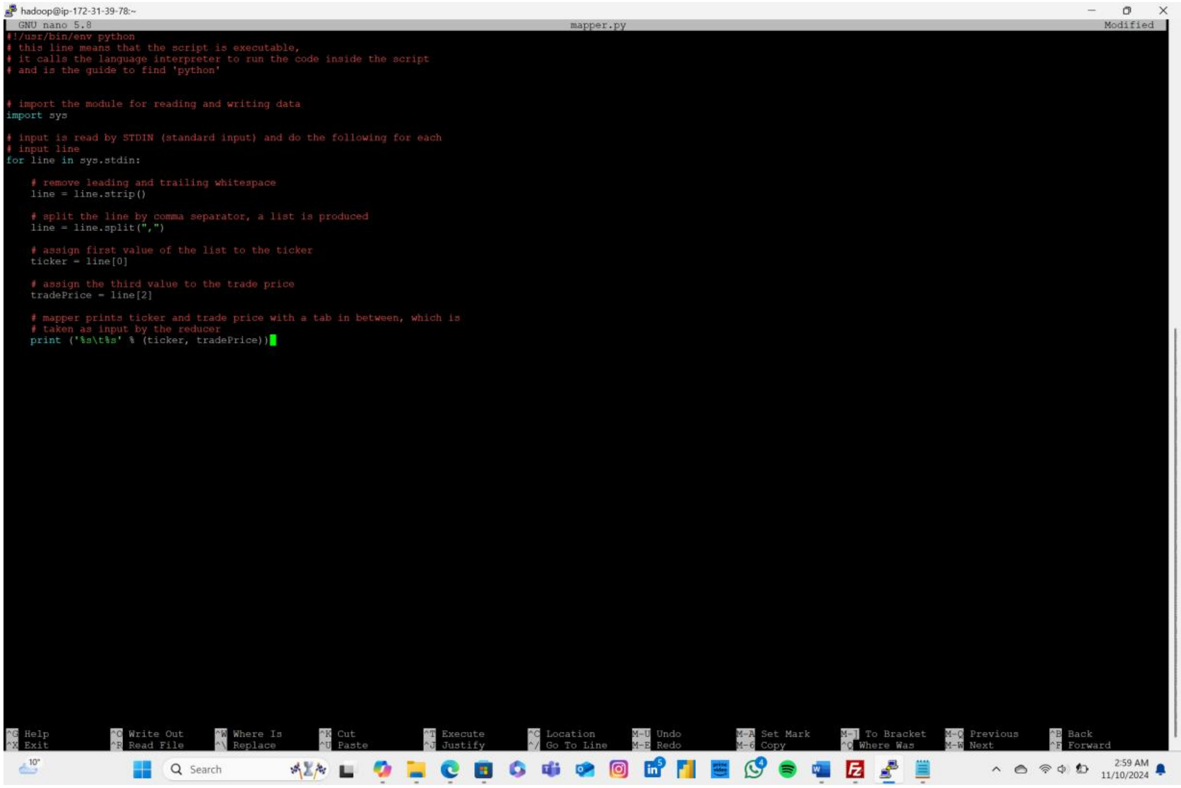
hadoop@ip-172-31-39-78:~

```
  GNU nano 5.8                                  NYSE_DATA.txt
CAH,Mon,48
GE,Tue,12
MCD,Wed,170
PFE,Thu,40
BAC,Fri,26
CAH,Mon,55
GE,Tue,13
MCD,Wed,160
PFE,Thu,42
BAC,Fri,25
CAH,Mon,56
GE,Tue,7
MCD,Wed,162
PFE,Thu,38
BAC,Fri,21
CAH,Mon,50
GE,Tue,15
MCD,Wed,140
PFE,Thu,45
BAC,Fri,23
CAH,Mon,42
GE,Tue,13
MCD,Wed,175
PFE,Thu,44
BAC,Fri,25

                                  [ Read 25 lines (Converted from DOS format) ]
^G Help       ^O Write Out  ^W Where Is   ^K Cut        ^T Execute    ^C Location   M-U Undo      M-A Set Mark  M-] To Bracket M-Q Previous  ^B Back
^X Exit       ^R Read File  ^\ Replace    ^U Paste      ^J Justify    ^/ Go To Line M-E Redo      M-6 Copy      ^Q Where Was   M-W Next      ^F Forward
```



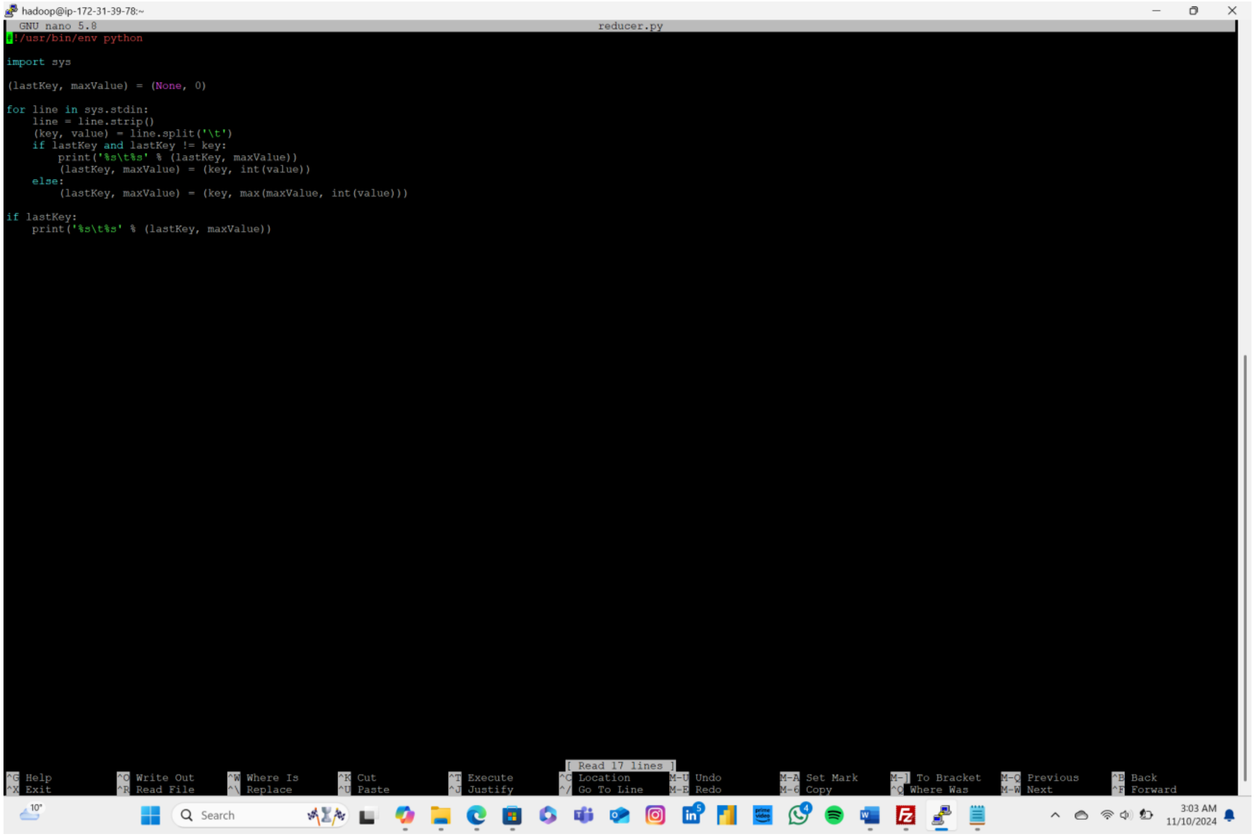
```
hadoop@ip-172-31-39-78:~
 Using username "hadoop".
 Authenticating with public key "My Key 11-10-2024"

A newer release of "Amazon Linux" is available.
  Version 2023.6.20241028:
  Version 2023.6.20241031:
Run "/usr/bin/dnf check-release-update" for full release and version update info
   ,     #_
   ~\_  ####_        Amazon Linux 2023
  ~~  \_#####\
  ~~     \###|
  ~~       \#/ ___   https://aws.amazon.com/linux/amazon-linux-2023
   ~~       V~' '->
    ~~~         /
      ~~._.   _/
         _/ _/
       _/m/'
Last login: Sun Nov 10 08:35:33 2024

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 MMMMMMM             MMMMMMM RRRRRRR      RRRRRR

[hadoop@ip-172-31-39-78 ~]$ ls
NYSE_DATA.txt
[hadoop@ip-172-31-39-78 ~]$ nano NYSE_*
[hadoop@ip-172-31-39-78 ~]$ nano NYSE_*
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -ls
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -mkdir MR
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -mkdir MR
mkdir: `MR': File exists
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -ls MR
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -copyFromLocal NYSE_DATA.txt MR
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -ls MR
Found 1 items
-rw-r--r--   1 hadoop hdfsadmingroup        297 2024-11-10 08:56 MR/NYSE_DATA.txt
[hadoop@ip-172-31-39-78 ~]$ 
```

4. Use nano to create two files mapper.py and reducer.py



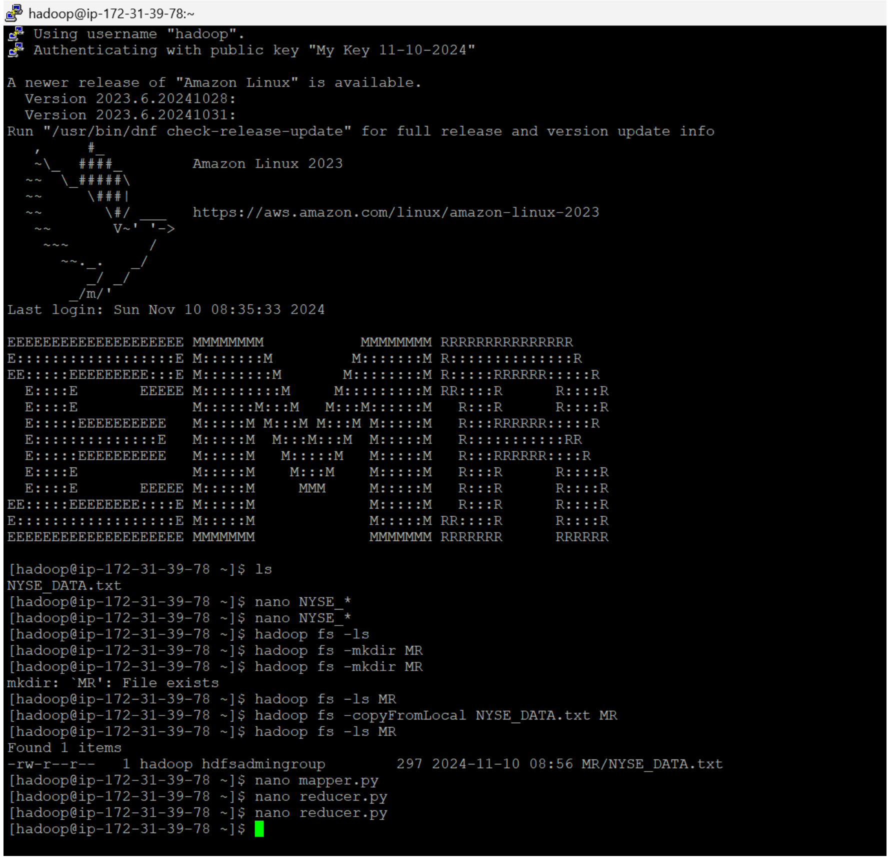
```
hadoop@ip-172-31-39-78:~
  GNU nano 5.8                         mapper.py                         Modified
#!/usr/bin/env python
# this line means that the script is executable,
# it calls the language interpreter to run the code inside the script
# and is the guide to find 'python'


# import the module for reading and writing data
import sys

# input is read by STDIN (standard input) and do the following for each
# input line
for line in sys.stdin:

    # remove leading and trailing whitespace
    line = line.strip()

    # split the line by comma separator, a list is produced
    line = line.split(",")

    # assign first value of the list to the ticker
    ticker = line[0]

    # assign the third value to the trade price
    tradePrice = line[2]

    # mapper prints ticker and trade price with a tab in between, which is
    # taken as input by the reducer
    print ('%s\t%s' % (ticker, tradePrice))
```

```
^G Help       ^O Write Out  ^W Where Is   ^K Cut        ^T Execute    ^C Location   M-U Undo      M-A Set Mark  M-] To Bracket  M-Q Previous  ^B Back
^X Exit       ^R Read File  ^\ Replace    ^U Paste      ^J Justify    ^/ Go To Line M-E Redo      M-6 Copy      ^Q Where Was    M-W Next      ^F Forward
```

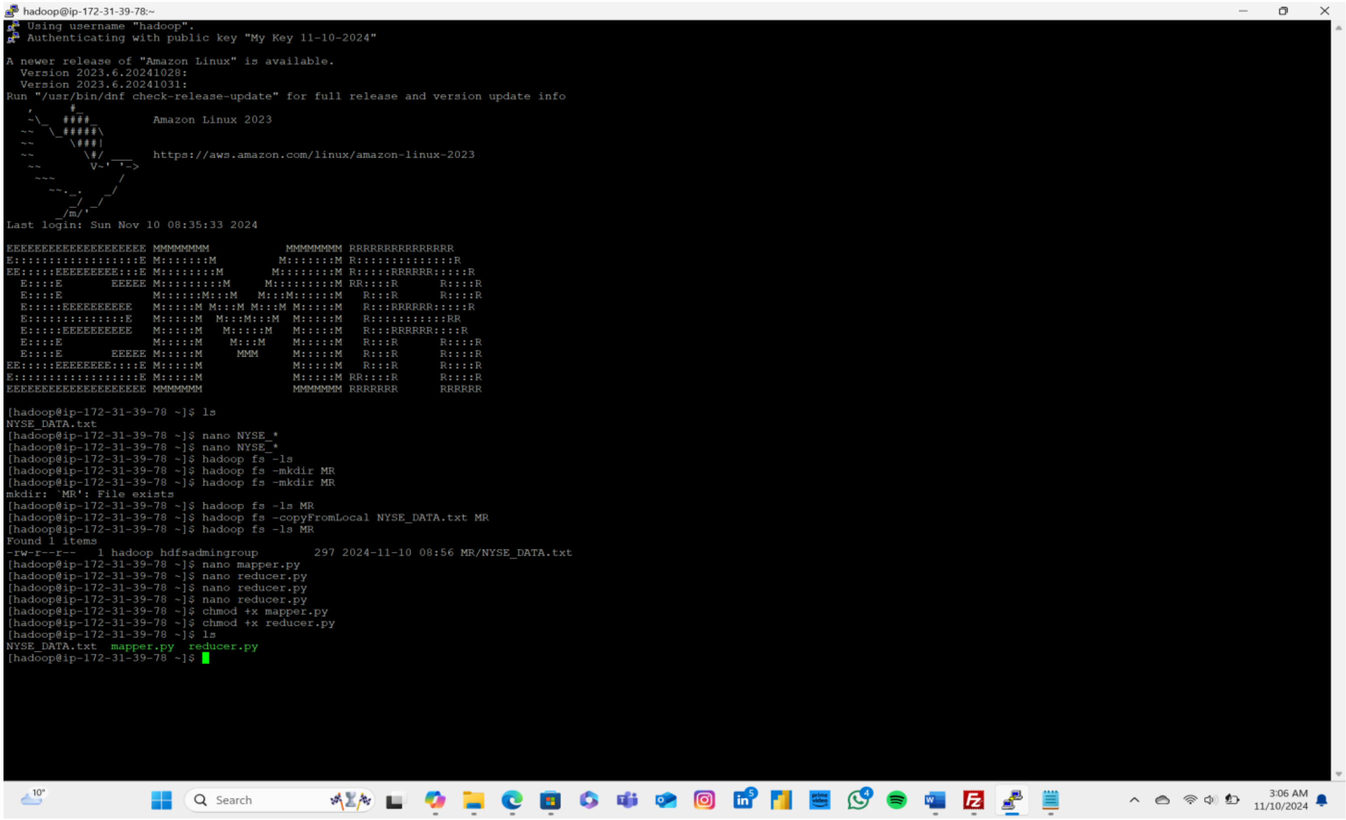


```
#!/usr/bin/env python

import sys

(lastKey, maxValue) = (None, 0)

for line in sys.stdin:
    line = line.strip()
    (key, value) = line.split('\t')
    if lastKey and lastKey != key:
        print('%s\t%s' % (lastKey, maxValue))
        (lastKey, maxValue) = (key, int(value))
    else:
        (lastKey, maxValue) = (key, max(maxValue, int(value)))

if lastKey:
    print('%s\t%s' % (lastKey, maxValue))
```

```
[ Read 17 lines ]
```

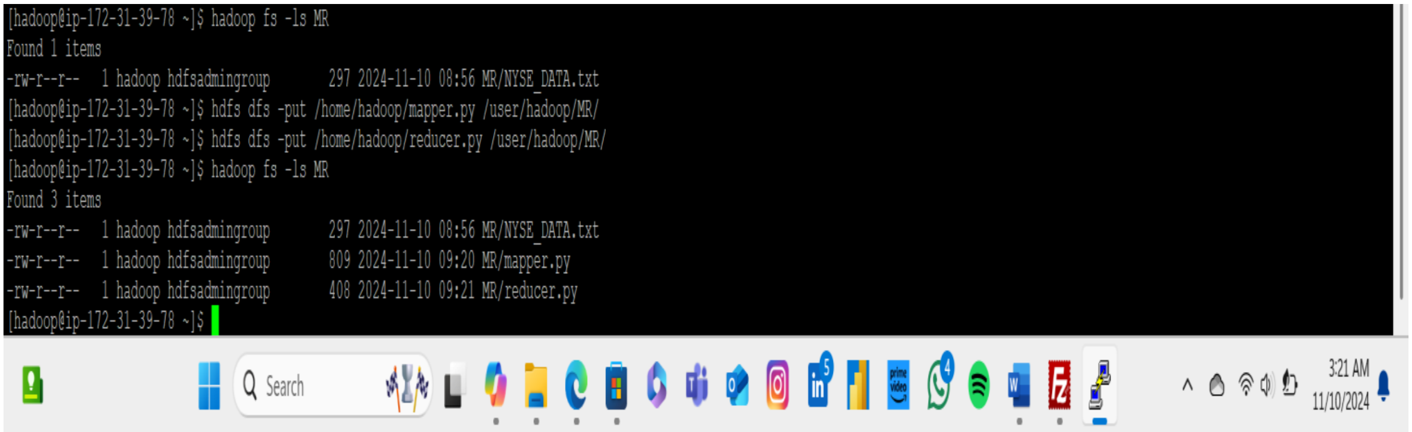


```
 Using username "hadoop".
 Authenticating with public key "My Key 11-10-2024"

A newer release of "Amazon Linux" is available.
  Version 2023.6.20241028:
  Version 2023.6.20241031:
Run "/usr/bin/dnf check-release-update" for full release and version update info
   ,     #_
   ~\_  ####_        Amazon Linux 2023
  ~~  \_#####\
  ~~     \###|
  ~~       \#/ ___   https://aws.amazon.com/linux/amazon-linux-2023
   ~~       V~' '->
    ~~~         /
      ~~._.   _/
         _/ _/
       _/m/'
Last login: Sun Nov 10 08:35:33 2024

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 MMMMMMM             MMMMMMM RRRRRRR      RRRRRR

[hadoop@ip-172-31-39-78 ~]$ ls
NYSE_DATA.txt
[hadoop@ip-172-31-39-78 ~]$ nano NYSE_*
[hadoop@ip-172-31-39-78 ~]$ nano NYSE_*
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -ls
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -mkdir MR
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -mkdir MR
mkdir: `MR': File exists
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -ls MR
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -copyFromLocal NYSE_DATA.txt MR
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -ls MR
Found 1 items
-rw-r--r--   1 hadoop hdfsadmingroup        297 2024-11-10 08:56 MR/NYSE_DATA.txt
[hadoop@ip-172-31-39-78 ~]$ nano mapper.py
[hadoop@ip-172-31-39-78 ~]$ nano reducer.py
[hadoop@ip-172-31-39-78 ~]$ nano reducer.py
[hadoop@ip-172-31-39-78 ~]$ 
```



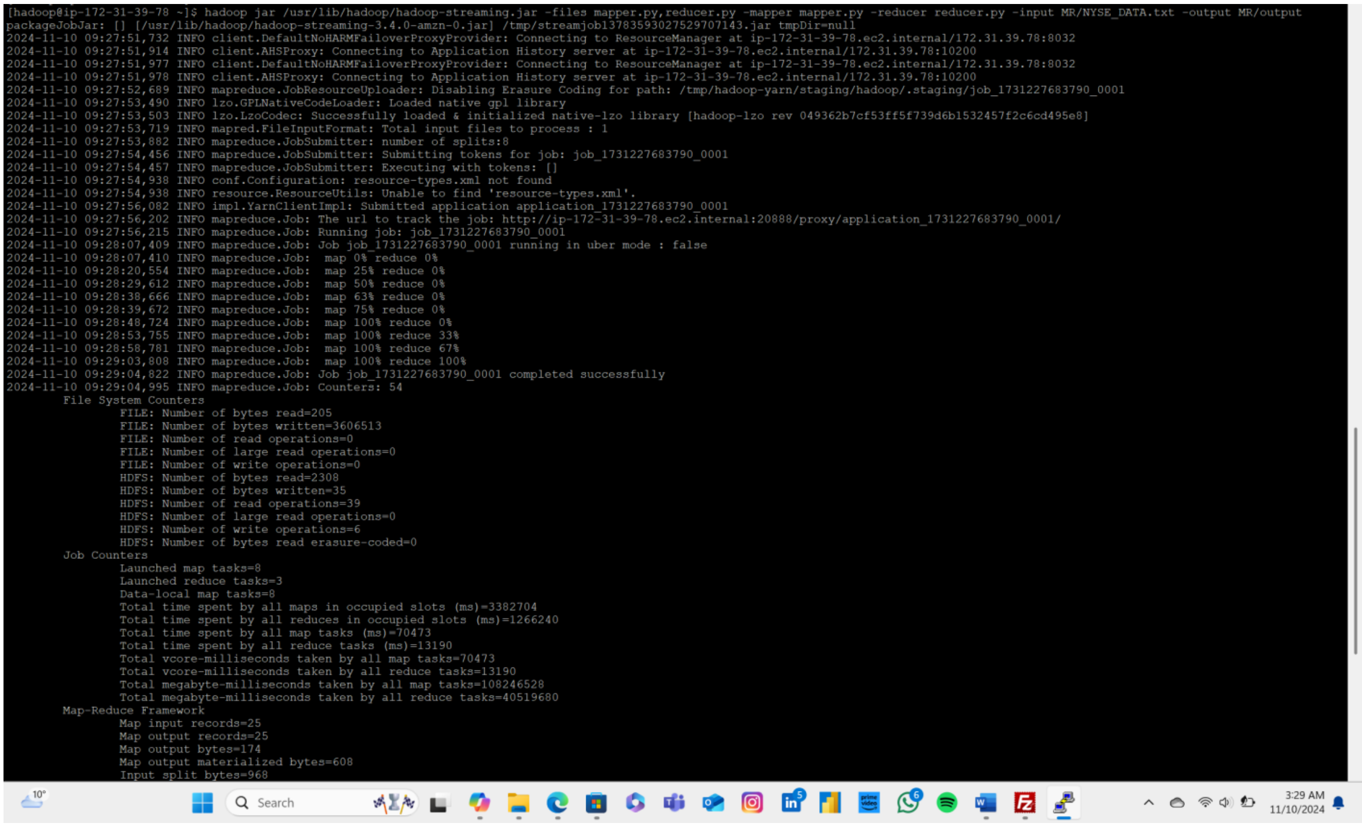
```
Using username "hadoop".
Authenticating with public key "My Key 11-10-2024"

A newer release of "Amazon Linux" is available.
  Version 2023.6.20241028:
  Version 2023.6.20241031:
Run "/usr/bin/dnf check-release-update" for full release and version update info
   ,     #_
   ~\_  ####_        Amazon Linux 2023
  ~~  \_#####\
  ~~     \###|
  ~~       \#/ ___   https://aws.amazon.com/linux/amazon-linux-2023
   ~~       V~' '->
    ~~~         /
      ~~._.   _/
         _/ _/
       _/m/'
Last login: Sun Nov 10 08:35:33 2024

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 MMMMMMM             MMMMMMM RRRRRRR      RRRRRR

[hadoop@ip-172-31-39-78 ~]$ ls
NYSE_DATA.txt
[hadoop@ip-172-31-39-78 ~]$ nano NYSE_*
[hadoop@ip-172-31-39-78 ~]$ nano NYSE_*
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -ls
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -mkdir MR
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -mkdir MR
mkdir: `MR': File exists
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -ls MR
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -copyFromLocal NYSE_DATA.txt MR
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -ls MR
Found 1 items
-rw-r--r--   1 hadoop hdfsadmingroup        297 2024-11-10 08:56 MR/NYSE_DATA.txt
[hadoop@ip-172-31-39-78 ~]$ nano mapper.py
[hadoop@ip-172-31-39-78 ~]$ nano reducer.py
[hadoop@ip-172-31-39-78 ~]$ nano reducer.py
[hadoop@ip-172-31-39-78 ~]$ nano reducer.py
[hadoop@ip-172-31-39-78 ~]$ chmod +x mapper.py
[hadoop@ip-172-31-39-78 ~]$ chmod +x reducer.py
[hadoop@ip-172-31-39-78 ~]$ ls
NYSE_DATA.txt  mapper.py  reducer.py
[hadoop@ip-172-31-39-78 ~]$
```

5. Move two .py files to Hadoop fs under the directory MR

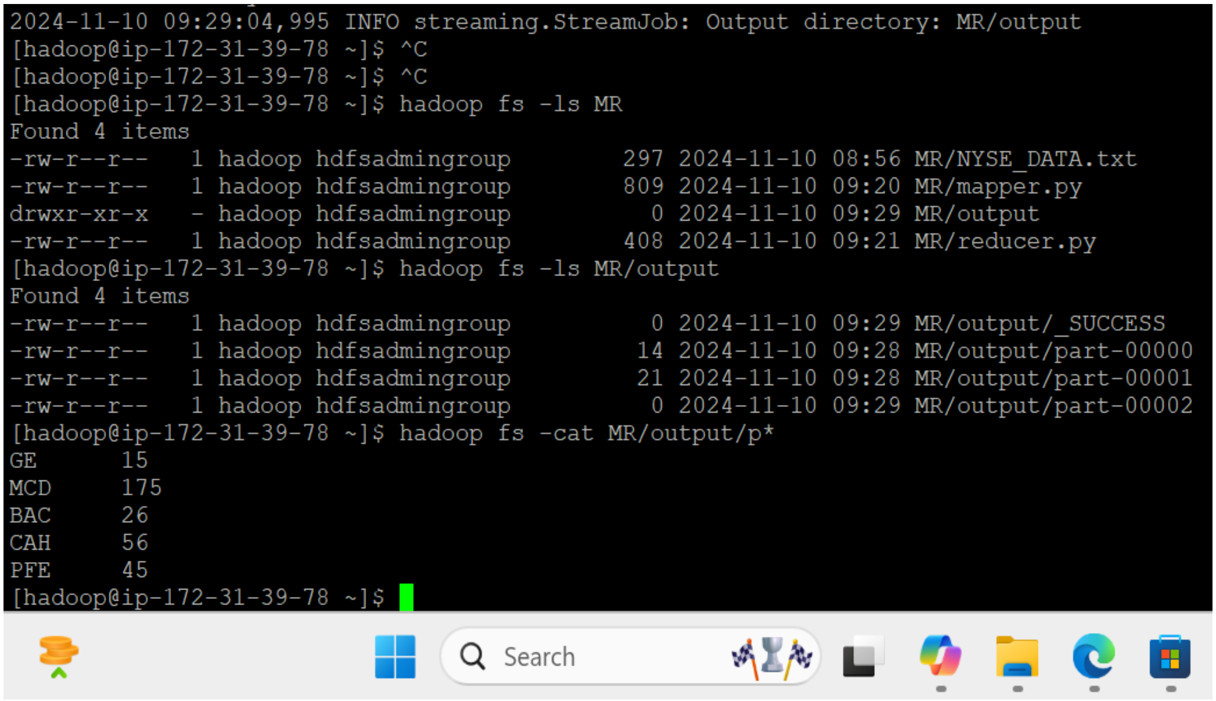


```
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -ls MR
Found 1 items
-rw-r--r--   1 hadoop hdfsadmingroup        297 2024-11-10 08:56 MR/NYSE_DATA.txt
[hadoop@ip-172-31-39-78 ~]$ hdfs dfs -put /home/hadoop/mapper.py /user/hadoop/MR/
[hadoop@ip-172-31-39-78 ~]$ hdfs dfs -put /home/hadoop/reducer.py /user/hadoop/MR/
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -ls MR
Found 3 items
-rw-r--r--   1 hadoop hdfsadmingroup        297 2024-11-10 08:56 MR/NYSE_DATA.txt
-rw-r--r--   1 hadoop hdfsadmingroup        809 2024-11-10 09:20 MR/mapper.py
-rw-r--r--   1 hadoop hdfsadmingroup        408 2024-11-10 09:21 MR/reducer.py
[hadoop@ip-172-31-39-78 ~]$
```

6. Run the MapReduce job command (remember to copy and paste the command from the MapReduce command file and -cat the output



```
[hadoop@ip-172-31-39-78 ~]$ hadoop jar /usr/lib/hadoop/hadoop-streaming.jar -files mapper.py,reducer.py -mapper mapper.py -reducer reducer.py -input MR/NYSE_DATA.txt -output MR/output
packageJobJar: [] [/usr/lib/hadoop/hadoop-streaming-3.4.0-amzn-0.jar] /tmp/streamjob13783593027529707143.jar tmpDir=null
2024-11-10 09:27:51,732 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at ip-172-31-39-78.ec2.internal/172.31.39.78:8032
2024-11-10 09:27:51,914 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-39-78.ec2.internal/172.31.39.78:10200
2024-11-10 09:27:51,977 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at ip-172-31-39-78.ec2.internal/172.31.39.78:8032
2024-11-10 09:27:51,978 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-39-78.ec2.internal/172.31.39.78:10200
2024-11-10 09:27:52,689 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1731227683790_0001
2024-11-10 09:27:53,490 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library
2024-11-10 09:27:53,503 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo library [hadoop-lzo rev 049362b7cf53ff5f739d6b1532457f2c6cd495e8]
2024-11-10 09:27:53,719 INFO mapred.FileInputFormat: Total input files to process : 1
2024-11-10 09:27:53,882 INFO mapreduce.JobSubmitter: number of splits:8
2024-11-10 09:27:54,456 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1731227683790_0001
2024-11-10 09:27:54,457 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-11-10 09:27:54,938 INFO conf.Configuration: resource-types.xml not found
2024-11-10 09:27:54,938 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-11-10 09:27:56,082 INFO impl.YarnClientImpl: Submitted application application_1731227683790_0001
2024-11-10 09:27:56,202 INFO mapreduce.Job: The url to track the job: http://ip-172-31-39-78.ec2.internal:20888/proxy/application_1731227683790_0001/
2024-11-10 09:27:56,215 INFO mapreduce.Job: Running job: job_1731227683790_0001
2024-11-10 09:28:07,409 INFO mapreduce.Job: Job job_1731227683790_0001 running in uber mode : false
2024-11-10 09:28:07,410 INFO mapreduce.Job:  map 0% reduce 0%
2024-11-10 09:28:20,554 INFO mapreduce.Job:  map 25% reduce 0%
2024-11-10 09:28:29,612 INFO mapreduce.Job:  map 50% reduce 0%
2024-11-10 09:28:38,666 INFO mapreduce.Job:  map 63% reduce 0%
2024-11-10 09:28:39,672 INFO mapreduce.Job:  map 75% reduce 0%
2024-11-10 09:28:48,724 INFO mapreduce.Job:  map 100% reduce 0%
2024-11-10 09:28:53,755 INFO mapreduce.Job:  map 100% reduce 33%
2024-11-10 09:28:58,781 INFO mapreduce.Job:  map 100% reduce 67%
2024-11-10 09:29:03,808 INFO mapreduce.Job:  map 100% reduce 100%
2024-11-10 09:29:04,822 INFO mapreduce.Job: Job job_1731227683790_0001 completed successfully
2024-11-10 09:29:04,995 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=205
                FILE: Number of bytes written=3606513
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=2308
                HDFS: Number of bytes written=35
                HDFS: Number of read operations=39
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=6
                HDFS: Number of bytes read erasure-coded=0
        Job Counters
                Launched map tasks=8
                Launched reduce tasks=3
                Data-local map tasks=8
                Total time spent by all maps in occupied slots (ms)=3382704
                Total time spent by all reduces in occupied slots (ms)=1266240
                Total time spent by all map tasks (ms)=70473
                Total time spent by all reduce tasks (ms)=13190
                Total vcore-milliseconds taken by all map tasks=70473
                Total vcore-milliseconds taken by all reduce tasks=13190
                Total megabyte-milliseconds taken by all map tasks=108246528
                Total megabyte-milliseconds taken by all reduce tasks=40519680
        Map-Reduce Framework
                Map input records=25
                Map output records=25
                Map output bytes=174
                Map output materialized bytes=608
                Input split bytes=968
```



```
2024-11-10 09:29:04,995 INFO streaming.StreamJob: Output directory: MR/output
[hadoop@ip-172-31-39-78 ~]$ ^C
[hadoop@ip-172-31-39-78 ~]$ ^C
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -ls MR
Found 4 items
-rw-r--r--   1 hadoop hdfsadmingroup        297 2024-11-10 08:56 MR/NYSE_DATA.txt
-rw-r--r--   1 hadoop hdfsadmingroup        809 2024-11-10 09:20 MR/mapper.py
drwxr-xr-x   - hadoop hdfsadmingroup          0 2024-11-10 09:29 MR/output
-rw-r--r--   1 hadoop hdfsadmingroup        408 2024-11-10 09:21 MR/reducer.py
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -ls MR/output
Found 4 items
-rw-r--r--   1 hadoop hdfsadmingroup          0 2024-11-10 09:29 MR/output/_SUCCESS
-rw-r--r--   1 hadoop hdfsadmingroup         14 2024-11-10 09:28 MR/output/part-00000
-rw-r--r--   1 hadoop hdfsadmingroup         21 2024-11-10 09:28 MR/output/part-00001
-rw-r--r--   1 hadoop hdfsadmingroup          0 2024-11-10 09:29 MR/output/part-00002
[hadoop@ip-172-31-39-78 ~]$ hadoop fs -cat MR/output/p*
GE      15
MCD     175
BAC     26
CAH     56
PFE     45
[hadoop@ip-172-31-39-78 ~]$
```

1. What the mapper.py does?

Role of Mapper: The mapper is responsible for processing the input data and converting it into key-value pairs.
In this script, each line of the input data is read, trimmed of whitespace, and split by a comma.
From the split data, the first element (representing the stock symbol) is chosen as the key, and the third element (representing the trade price) is chosen as the value.
The script outputs each stock symbol and its corresponding trade price, separated by a tab.

2. What the reducer.py does?
   Role of Reducer: The reducer aggregates the key-value pairs output by the mapper.
   This script reads the mapper's output, splits each line by the tab to separate the key and value, and keeps track of the maximum trade price for each stock symbol.
   If the current key is the same as the last key, it compares the current trade price with the maximum trade price seen so far and updates it if necessary. If the current key is different, it outputs the last key and its maximum trade price, and resets for the new key.
   After processing all lines, the script outputs the maximum trade price for each stock symbol.

3. Output:
   Structure of Output Directory: The output directory contains the results of the MapReduce job. It includes files named part-00000, part-00001, etc., each containing a portion of the result, and a _SUCCESS file indicating successful job completion.
   The output files contain key-value pairs, where each key is a stock symbol, and each value is the maximum trade price for that stock symbol.

   Example Output Content:

   ```
   GE     15
   MCD    175
   BAC    26
   CAH    56
   PFE    45
   ```
   Each line contains a stock symbol and its maximum trade price, separated by a tab.