Program 4: Implement matrix multiplication with Hadoop Map Reduce.

: Login into Hadoop user you used while installing Hadoop , here we use hdoop user

```
hdoop@NuvobookV1:~$
```

$ su – hdoop

-start the Hadoop server

$ cd hadoop-3.4.0/sbin/

$ ./start-dfs.sh

$ ./start-yarn

$ jps

```
hdoop@hdoop-VirtualBox:~$ cd hadoop-3.2.1/sbin
hdoop@hdoop-VirtualBox:~/hadoop-3.2.1/sbin$ ./start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [hdoop-VirtualBox]
hdoop@hdoop-VirtualBox:~/hadoop-3.2.1/sbin$ ./start-yarn.sh
Starting resourcemanager
Starting nodemanagers
hdoop@hdoop-VirtualBox:~/hadoop-3.2.1/sbin$ jps
4817 DataNode
5298 ResourceManager
5000 SecondaryNameNode
5450 NodeManager
4683 NameNode
5982 Jps
hdoop@hdoop-VirtualBox:~/hadoop-3.2.1/sbin$
```

#make a dir for the program 4

$ hdfs dfs -mkdir /p4

#create mapper program

$sudo nano mapper.py

-insert this code

```
#!/usr/bin/env python3

import sys

for line in sys.stdin:
    line = line.strip()
    entry = line.split(',')
    key=entry[0]
    value = line[1:]
    if key =='A':
        print('{0}\t{1}'.format(key,value))
    elif key =='B':
        print('{0}\t{1}'.format(key,value))
```

-click Ctrl+s and Ctrl+x to close it

note: make sure you have python3 in your system

#reducer program

$sudo nano reducer.py

-insert this code

```
#!/usr/bin/env python
import sys
mat1 = {}
mat2 = {}

# Reading input and populating the matrices
for line in sys.stdin:
  line = line.strip()
  key, value = line.split('\t')
  v = value.split(',')
  if key == 'A':
    mat1[(int(v[1]), int(v[2]))] = int(v[3])
  elif key == 'B':  # Corrected from '8' to 'B'
    mat2[(int(v[1]), int(v[2]))] = int(v[3])

result = 0
# Multiplying the matrices with safe access to the elements
for i in range(0, 3):
  for j in range(0, 3):
    for k in range(0, 3):
      # Using .get() method to avoid KeyError
      result += mat1.get((i, k), 0) * mat2.get((k, j), 0)
    print('({0},{1})\t{2}'.format(i, j, result))
```

result = 0

#create the matrix file :

$ sudo nano matrixinput.txt

-insert the data

A,0,0,1

A,0,1,2

A,0,2,3

A,1,0,4

A,1,1,5

A,1,2,6

A,2,0,7

A,2,1,8

A,2,2,9

B,0,0,1

B,0,1,1

B,0,2,1

B,1,0,1

B,1,1,2

B,1,2,3

B,2,0,1

B,2,1,1

B,2,2,1

#change the permissions of the files

$sudo chmod 777 mapper.py reducer.py

#upload the dataset to Hadoop

$ hdfs dfs -copyFromLocal matrixinput.txt /p4/

```
hdoop@NuvobookV1:~/lab/p4$ hdfs dfs -copyFromLocal matrixinput.txt /p4/
hdoop@NuvobookV1:~/lab/p4$ hdfs dfs -ls /p4
Found 1 items
-rw-r--r--   1 hdoop supergroup        145 2024-10-17 07:12 /p4/matrixinput.
txt
hdoop@NuvobookV1:~/lab/p4$
```

#running python program in Hadoop using Hadoop streamer

$ cd

$ wget https://repo1.maven.org/maven2/org/apache/hadoop/hadoopstreaming/2.7.3/hadoop-streaming-2.7.3.jar $ hadoop jar /home/hdoop/hadoop-streaming-2.7.3.jar -input /p2/sample.txt -output /p2/output -mapper /home/hdoop/p2/mapper.py -reducer /home/hdoop/p2/reducer.py

$ hadoop jar /home/hdoop/hadoop-streaming-2.7.3.jar -input /p4/matrixinput.txt -output /p4/output -mapper /home/hdoop/mapper.py -reducer /home/hdoop/reducer.py

#output

```
hdoop@NuvobookV1:~/lab/p4$ hadoop jar /home/hdoop/hadoop-streaming-2.7.3.jar -input /p4/matrixinput.txt -output /p4/output -mapper /home/hdoop/lab/p4/mapper
.py -reducer /home/hdoop/lab/p4/reducer.py
packageJobJar: [/tmp/hadoop-unjar6029587034847901962/] [] /tmp/streamjob6834679690450673703.jar tmpDir=null
2024-10-17 07:13:55,545 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /172.18.132.51:8032
2024-10-17 07:13:55,839 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /172.18.132.51:8032
2024-10-17 07:13:56,221 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_1729128507769_000
9
2024-10-17 07:13:56,504 INFO mapred.FileInputFormat: Total input files to process : 1
2024-10-17 07:13:56,585 INFO mapreduce.JobSubmitter: number of splits:2
2024-10-17 07:13:56,788 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1729128507769_0009
2024-10-17 07:13:56,789 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-10-17 07:13:57,003 INFO conf.Configuration: resource-types.xml not found
2024-10-17 07:13:57,004 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-10-17 07:13:57,431 INFO impl.YarnClientImpl: Submitted application application_1729128507769_0009
2024-10-17 07:13:57,471 INFO mapreduce.Job: The url to track the job: http://172.18.132.51:8088/proxy/application_1729128507769_0009/
2024-10-17 07:13:57,473 INFO mapreduce.Job: Running job: job_1729128507769_0009
2024-10-17 07:14:04,625 INFO mapreduce.Job: Job job_1729128507769_0009 running in uber mode : false
2024-10-17 07:14:04,626 INFO mapreduce.Job:  map 0% reduce 0%
2024-10-17 07:14:11,870 INFO mapreduce.Job:  map 100% reduce 0%
2024-10-17 07:14:16,917 INFO mapreduce.Job:  map 100% reduce 100%
2024-10-17 07:14:18,955 INFO mapreduce.Job: Job job_1729128507769_0009 completed successfully
2024-10-17 07:14:19,192 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=204
                FILE: Number of bytes written=934346
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=410
                HDFS: Number of bytes written=79
                HDFS: Number of read operations=11
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
        Job Counters
                Launched map tasks=2
                Launched reduce tasks=1
                Data-local map tasks=2
                Total time spent by all maps in occupied slots (ms)=9464
                Total time spent by all reduces in occupied slots (ms)=2657
                Total time spent by all map tasks (ms)=9464
                Total time spent by all reduce tasks (ms)=2657
                Total vcore-milliseconds taken by all map tasks=9464
                Total vcore-milliseconds taken by all reduce tasks=2657
                Total vcore-milliseconds taken by all map tasks=9464
                Total vcore-milliseconds taken by all reduce tasks=2657
                Total megabyte-milliseconds taken by all map tasks=9691136
                Total megabyte-milliseconds taken by all reduce tasks=2720768
        Map-Reduce Framework
                Map input records=19
                Map output records=18
                Map output bytes=162
                Map output materialized bytes=210
                Input split bytes=192
                Combine input records=0
                Combine output records=0
                Reduce input groups=2
                Reduce shuffle bytes=210
                Reduce input records=18
                Reduce output records=9
                Spilled Records=36
                Shuffled Maps =2
                Failed Shuffles=0
                Merged Map outputs=2
                GC time elapsed (ms)=183
                CPU time spent (ms)=3290
                Physical memory (bytes) snapshot=878444544
                Virtual memory (bytes) snapshot=7690653696
                Total committed heap usage (bytes)=668991488
                Peak Map Physical memory (bytes)=320208896
                Peak Map Virtual memory (bytes)=2560774144
                Peak Reduce Physical memory (bytes)=238653440
                Peak Reduce Virtual memory (bytes)=2570227712
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=218
        File Output Format Counters
                Bytes Written=79
2024-10-17 07:14:19,192 INFO streaming.StreamJob: Output directory: /p4/output
hdoop@NuvobookV1:~/lab/p4$
```

$ hdfs dfs -cat /p4/output/part-00000

```
hdoop@NuvobookV1:~/lab/p4$ hdfs dfs -cat /p4/output/part-00000
(0,0)   6
(0,1)   8
(0,2)   10
(1,0)   15
(1,1)   20
(1,2)   25
(2,0)   24
(2,1)   32
(2,2)   40
hdoop@NuvobookV1:~/lab/p4$
```