VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Big Data Analytics Lab
PMDS507P

Name: **Tufan Kundu** Registration number: **24MDT0184**

Slot: L29+L30

Digital Assignment 5

Explain the steps involved in performing a Matrix multiplication in Hadoop

Step 1: Preparing the A matrix in A.txt

input > A.txt

```
A 0 0 1
A 0 1 2
A 1 0 3
A 1 1 4
```

Step 2: Preparing the B matrix B.txt

input > B.txt

```
B 0 0 5
B 0 1 6
B 1 0 7
B 1 1 8
```

Step 3: Mapper.py

mapper.py > ...

```python
#!/usr/bin/env python3
import sys

for line in sys.stdin:
    line = line.strip()
    if not line:
        continue
    matrix, row, col, value = line.split()
    row = int(row)
    col = int(col)
    value = float(value)

    n = 2

    if matrix == 'A':
        for k in range(n):
            print(f"{row},{k}\tA, {col}, {value}")
    elif matrix == "B":
        for i in range(n):
            print(f"{i},{col}\tB,{row},{value}")

```

Step 4: Reducer.py

reducer.py > ...

```python
#!/usr/bin/env python3
import sys
from collections import defaultdict

current_key = None
a_values = defaultdict(float)
b_values = defaultdict(float)

for line in sys.stdin:
    line = line.strip()
    key, value = line.split('\t')
    if key!= current_key:
        if current_key:
            result = 0
            for j in a_values:
                result += a_values[j]*b_values[j]
            print(f"{current_key}\t{result}")
        current_key = key
        a_values.clear()
        b_values.clear()
    tag, index, val = value.split(',')
    index = int(index)
    val = float(val)

    if tag == 'A':
        a_values[index] = val
    else:
        b_values[index] = val

if current_key:
    result = 0
    for j in a_values:
        result+= a_values[j]*b_values[j]
    print(f"{current_key}\t{result}")
```

Step 5: Start HDFS and YARN and verify with jps

```
hduser@sjt217score015:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hduser in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [sjt217score015]
2025-10-17 13:00:40,014 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting resourcemanager
Starting nodemanagers
hduser@sjt217score015:~$ jps
91458 SecondaryNameNode
90951 NameNode
91769 ResourceManager
92377 Jps
91178 DataNode
92141 NodeManager
```

Step 6: Making Scripts executable

chmod +x /home/hduser/Desktop/operationhadoop/mapper.py
chmod +x /home/hduser/Desktop/operationhadoop/reducer.py

```
hduser@sjt217score015:~$ chmod +x /home/hduser/Desktop/matrix_multiplication/mapper.py
hduser@sjt217score015:~$ chmod +x /home/hduser/Desktop/matrix_multiplication/reducer.py
```

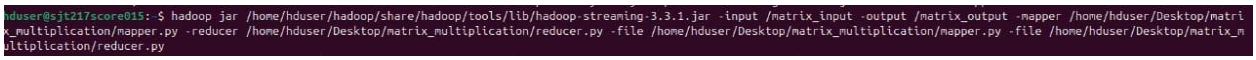
Step 7: Creating directory

hdfs dfs -mkdir /matrix_input
hdfs dfs -put /home/hduser/Desktop/matrix_multiplication/input/* /matrix_input/

```
hduser@sjt217score015:~$ hdfs dfs -mkdir /matrix_input
2025-10-17 13:01:47,447 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduser@sjt217score015:~$ hdfs dfs -put /home/hduser/Desktop/matrix_multiplication/input/* /matrix_input/
2025-10-17 13:02:21,487 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```

Step 8: Running the mapper and reducer

hadoop jar /home/hduser/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.3.1.jar -input /matrix_input -output /matrix_output -mapper /home/hduser/Desktop/matrix_multiplication/mapper.py -reducer /home/hduser/Desktop/matrix_multiplication/reducer.py -file /home/hduser/Desktop/matrix_multiplication/mapper.py -file /home/hduser/Desktop/matrix_multiplication/reducer.py



Step 9: Displaying the result

hdfs dfs -cat /matrix_output/part-00000