EXP NO: 3

DATE:

Implement of Matrix Multiplication with Hadoop Map Reduce

AIM: -

BACKGROUND THEORY: -

PROCEDURE: -

- Switch to superuser mode using sudo su.
- Input Format:
  - We'll have two input files representing matrices A and B.
  - Matrix A (m x n): Split by rows.
  - Matrix B (n x p): Split by columns.
- Map Step:
  - For each element in matrix A (i, k, A[i][k]), emit the product to the intermediate key (i, j) where j is the column index in matrix B.
  - For each element in matrix B (k, j, B[k][j]), emit the product to the intermediate key (i, j) for all rows i of matrix A.
- Reduce Step:
  - For each intermediate key (i, j), sum the products of the corresponding values to calculate the result matrix C (i, j).
- Output Format:
  - Output matrix C, where each line is of the form i, j, C[i][j].

CODING: -

- sudo su
- MatrixMapper.java

```java
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

public class MatrixMapper extends Mapper<Object, Text, Text, Text> {

    @Override
    public void map(Object key, Text value, Context context) throws IOException, InterruptedException {
        String[] line = value.toString().split(",");
        String matrixName = line[0];  // A or B
        int i = Integer.parseInt(line[1]);
        int j = Integer.parseInt(line[2]);
        int valueOfElement = Integer.parseInt(line[3]);

        if (matrixName.equals("A")) {
            // Emit for all columns of B
            for (int k = 0; k < context.getConfiguration().getInt("p", 0); k++) {
                context.write(new Text(i + "," + k), new Text("A," + j + "," + valueOfElement));
            }
        } else {
            // Emit for all rows of A
            for (int k = 0; k < context.getConfiguration().getInt("m", 0); k++) {
                context.write(new Text(k + "," + j), new Text("B," + i + "," + valueOfElement));
            }
        }
    }
}
```

- MatrixReducer.java

```
import java.io.IOException;
import java.util.HashMap;
import java.util.Map;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

public class MatrixReducer extends Reducer<Text, Text, Text, IntWritable> {

    @Override
    public void reduce(Text key, Iterable<Text> values, Context context) throws IOException, InterruptedException {
        Map<Integer, Integer> mapA = new HashMap<>();
        Map<Integer, Integer> mapB = new HashMap<>();

        for (Text val : values) {
            String[] parts = val.toString().split(",");
            if (parts[0].equals("A")) {
                mapA.put(Integer.parseInt(parts[1]), Integer.parseInt(parts[2]));
            } else {
                mapB.put(Integer.parseInt(parts[1]), Integer.parseInt(parts[2]));
            }
        }

        int result = 0;
        for (Integer k : mapA.keySet()) {
            if (mapB.containsKey(k)) {
                result += mapA.get(k) * mapB.get(k);
            }
        }

        context.write(key, new IntWritable(result));
    }
}
```

- MatrixMultiplication.java

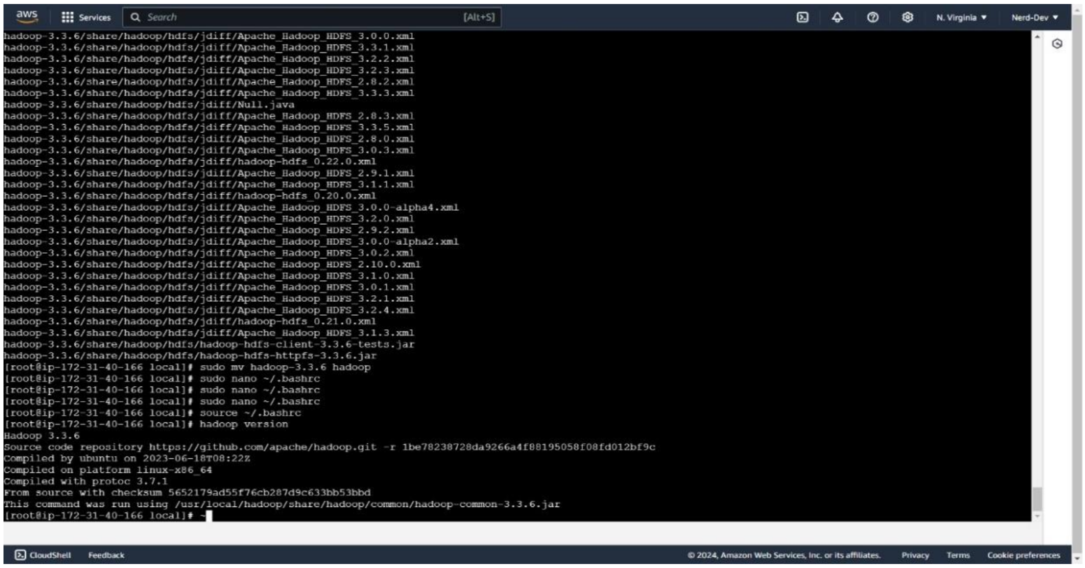
```
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
```

```
public class MatrixMultiplication {

    public static void main(String[] args) throws Exception {
        Configuration conf = new Configuration();
        // Dimensions of the matrices A (m x n) and B (n x p)
        conf.setInt("m", 3);  // Rows of A
        conf.setInt("n", 2);  // Columns of A and Rows of B
        conf.setInt("p", 3);  // Columns of B

        Job job = Job.getInstance(conf, "Matrix Multiplication");
        job.setJarByClass(MatrixMultiplication.class);
        job.setMapperClass(MatrixMapper.class);
        job.setReducerClass(MatrixReducer.class);

        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);

        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));
        System.exit(job.waitForCompletion(true) ? 0 : 1);
    }
}
```

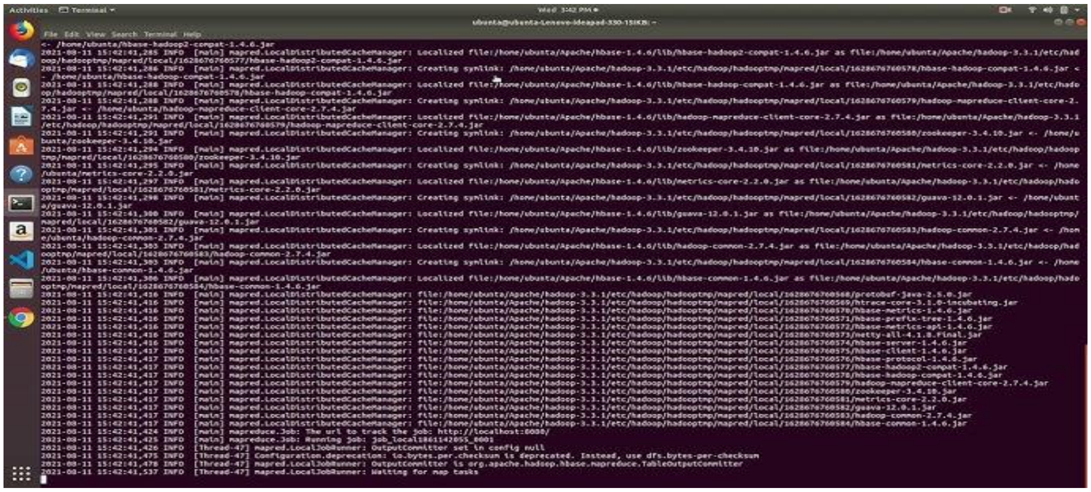
- Steps to Run the Code

```
hdfs dfs -put matrixA.txt /input/
hdfs dfs -put matrixB.txt /input/
hadoop com.sun.tools.javac.Main MatrixMultiplication.java
jar cf matrixmultiplication.jar MatrixMultiplication*.class
hadoop jar matrixmultiplication.jar MatrixMultiplication /input/ /output/
hdfs dfs -cat /output/part-r-00000
```

OUTPUT: -



```
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.0.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.3.1.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.2.2.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.2.3.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.8.2.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.3.3.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Null.java
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.8.3.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.3.5.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.8.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.0.3.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/hadoop-hdfs_0.22.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.9.1.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.1.1.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/hadoop-hdfs_0.20.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.0.0-alpha4.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.2.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.9.2.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.0.0-alpha2.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.0.2.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_2.10.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.1.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.0.1.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.2.1.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.2.4.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/hadoop-hdfs_0.21.0.xml
hadoop-3.3.6/share/hadoop/hdfs/jdiff/Apache_Hadoop_HDFS_3.1.3.xml
hadoop-3.3.6/share/hadoop/hdfs/hadoop-hdfs-client-3.3.6-tests.jar
hadoop-3.3.6/share/hadoop/hdfs/hadoop-hdfs-httpfs-3.3.6.jar
[root@ip-172-31-40-166 local]# sudo mv hadoop-3.3.6 hadoop
[root@ip-172-31-40-166 local]# sudo nano ~/.bashrc
[root@ip-172-31-40-166 local]# sudo nano ~/.bashrc
[root@ip-172-31-40-166 local]# sudo nano ~/.bashrc
[root@ip-172-31-40-166 local]# source ~/.bashrc
[root@ip-172-31-40-166 local]# hadoop version
Hadoop 3.3.6
Source code repository https://github.com/apache/hadoop.git -r 1be78238728da9266a4f88195058f08fd012bf9c
Compiled by ubuntu on 2023-06-18T08:22Z
Compiled on platform linux-x86_64
Compiled with protoc 3.7.1
From source with checksum 5652179ad55f76cb287d9c633bb53bbd
This command was run using /usr/local/hadoop/share/hadoop/common/hadoop-common-3.3.6.jar
[root@ip-172-31-40-166 local]# ~
```


Activities
Terminal ▾
Wed 3:42 PM
ubuntu@ubuntu-Lenovo-ideapad-330-15IKB: ~
File Edit View Search Terminal Help

```
<- /home/ubuntu/hbase-hadoop2-compat-1.4.6.jar
2021-08-11 15:42:41,285 INFO  [main] mapred.LocalDistributedCacheManager: Localized file:/home/ubuntu/Apache/hbase-1.4.6/lib/hbase-hadoop2-compat-1.4.6.jar as file:/home/ubuntu/Apache/hadoop-3.3.1/etc/had
oop/hadooptmp/mapred/local/1628676760577/hbase-hadoop2-compat-1.4.6.jar
2021-08-11 15:42:41,286 INFO  [main] mapred.LocalDistributedCacheManager: Creating symlink: /home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760578/hbase-hadoop-compat-1.4.6.jar <
- /home/ubuntu/hbase-hadoop-compat-1.4.6.jar
2021-08-11 15:42:41,288 INFO  [main] mapred.LocalDistributedCacheManager: Localized file:/home/ubuntu/Apache/hbase-1.4.6/lib/hbase-hadoop-compat-1.4.6.jar as file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hado
op/hadooptmp/mapred/local/1628676760578/hbase-hadoop-compat-1.4.6.jar
2021-08-11 15:42:41,288 INFO  [main] mapred.LocalDistributedCacheManager: Creating symlink: /home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760579/hadoop-mapreduce-client-core-2.
7.4.jar <- /home/ubuntu/hadoop-mapreduce-client-core-2.7.4.jar
2021-08-11 15:42:41,291 INFO  [main] mapred.LocalDistributedCacheManager: Localized file:/home/ubuntu/Apache/hbase-1.4.6/lib/hadoop-mapreduce-client-core-2.7.4.jar as file:/home/ubuntu/Apache/hadoop-3.3.1
/etc/hadoop/hadooptmp/mapred/local/1628676760579/hadoop-mapreduce-client-core-2.7.4.jar
2021-08-11 15:42:41,291 INFO  [main] mapred.LocalDistributedCacheManager: Creating symlink: /home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760580/zookeeper-3.4.10.jar <- /home/u
buntu/zookeeper-3.4.10.jar
2021-08-11 15:42:41,294 INFO  [main] mapred.LocalDistributedCacheManager: Localized file:/home/ubuntu/Apache/hbase-1.4.6/lib/zookeeper-3.4.10.jar as file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadoop
tmp/mapred/local/1628676760580/zookeeper-3.4.10.jar
2021-08-11 15:42:41,295 INFO  [main] mapred.LocalDistributedCacheManager: Creating symlink: /home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760581/metrics-core-2.2.0.jar <- /home
/ubuntu/metrics-core-2.2.0.jar
2021-08-11 15:42:41,297 INFO  [main] mapred.LocalDistributedCacheManager: Localized file:/home/ubuntu/Apache/hbase-1.4.6/lib/metrics-core-2.2.0.jar as file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hado
optmp/mapred/local/1628676760581/metrics-core-2.2.0.jar
2021-08-11 15:42:41,298 INFO  [main] mapred.LocalDistributedCacheManager: Creating symlink: /home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760582/guava-12.0.1.jar <- /home/ubunt
u/guava-12.0.1.jar
2021-08-11 15:42:41,300 INFO  [main] mapred.LocalDistributedCacheManager: Localized file:/home/ubuntu/Apache/hbase-1.4.6/lib/guava-12.0.1.jar as file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/
mapred/local/1628676760582/guava-12.0.1.jar
2021-08-11 15:42:41,301 INFO  [main] mapred.LocalDistributedCacheManager: Creating symlink: /home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760583/hadoop-common-2.7.4.jar <- /hom
e/ubuntu/hadoop-common-2.7.4.jar
2021-08-11 15:42:41,303 INFO  [main] mapred.LocalDistributedCacheManager: Localized file:/home/ubuntu/Apache/hbase-1.4.6/lib/hadoop-common-2.7.4.jar as file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/had
ooptmp/mapred/local/1628676760583/hadoop-common-2.7.4.jar
2021-08-11 15:42:41,303 INFO  [main] mapred.LocalDistributedCacheManager: Creating symlink: /home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760584/hbase-common-1.4.6.jar <- /home
/ubuntu/hbase-common-1.4.6.jar
2021-08-11 15:42:41,306 INFO  [main] mapred.LocalDistributedCacheManager: Localized file:/home/ubuntu/Apache/hbase-1.4.6/lib/hbase-common-1.4.6.jar as file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hado
optmp/mapred/local/1628676760584/hbase-common-1.4.6.jar
2021-08-11 15:42:41,416 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760568/protobuf-java-2.5.0.jar
2021-08-11 15:42:41,416 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760569/htrace-core-3.1.0-incubating.jar
2021-08-11 15:42:41,416 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760570/hbase-metrics-1.4.6.jar
2021-08-11 15:42:41,416 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760571/hbase-prefix-tree-1.4.6.jar
2021-08-11 15:42:41,416 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760572/hbase-metrics-api-1.4.6.jar
2021-08-11 15:42:41,416 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760573/netty-all-4.1.8.Final.jar
2021-08-11 15:42:41,416 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760574/hbase-server-1.4.6.jar
2021-08-11 15:42:41,417 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760575/hbase-client-1.4.6.jar
2021-08-11 15:42:41,417 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760576/hbase-protocol-1.4.6.jar
2021-08-11 15:42:41,417 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760577/hbase-hadoop2-compat-1.4.6.jar
2021-08-11 15:42:41,417 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760578/hbase-hadoop-compat-1.4.6.jar
2021-08-11 15:42:41,417 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760579/hadoop-mapreduce-client-core-2.7.4.jar
2021-08-11 15:42:41,417 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760580/zookeeper-3.4.10.jar
2021-08-11 15:42:41,417 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760581/metrics-core-2.2.0.jar
2021-08-11 15:42:41,417 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760582/guava-12.0.1.jar
2021-08-11 15:42:41,417 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760583/hadoop-common-2.7.4.jar
2021-08-11 15:42:41,417 INFO  [main] mapred.LocalDistributedCacheManager: file:/home/ubuntu/Apache/hadoop-3.3.1/etc/hadoop/hadooptmp/mapred/local/1628676760584/hbase-common-1.4.6.jar
2021-08-11 15:42:41,424 INFO  [main] mapreduce.Job: The url to track the job: http://localhost:8080/
2021-08-11 15:42:41,425 INFO  [main] mapreduce.Job: Running job: job_local1861142055_0001
2021-08-11 15:42:41,426 INFO  [Thread-47] mapred.LocalJobRunner: OutputCommitter set in config null
2021-08-11 15:42:41,475 INFO  [Thread-47] Configuration.deprecation: io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2021-08-11 15:42:41,476 INFO  [Thread-47] mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.hbase.mapreduce.TableOutputCommitter
2021-08-11 15:42:41,537 INFO  [Thread-47] mapred.LocalJobRunner: Waiting for map tasks
```