

MATRIX MULTIPLICATION

Map Function

Algorithm 1: The Map Function

```

1 for each element  $m_{ij}$  of  $M$  do
2   | produce  $(key, value)$  pairs as  $((i, k), (M, j, m_{ij}))$ , for  $k = 1, 2, 3, \dots$  up
   | to the number of columns of  $N$ 
3 for each element  $n_{jk}$  of  $N$  do
4   | produce  $(key, value)$  pairs as  $((i, k), (N, j, n_{jk}))$ , for  $i = 1, 2, 3, \dots$  up
   | to the number of rows of  $M$ 
5 return Set of  $(key, value)$  pairs that each key,  $(i, k)$ , has a list with
   values  $(M, j, m_{ij})$  and  $(N, j, n_{jk})$  for all possible values of  $j$ 

```

Reduce Function

Algorithm 2: The Reduce Function

```

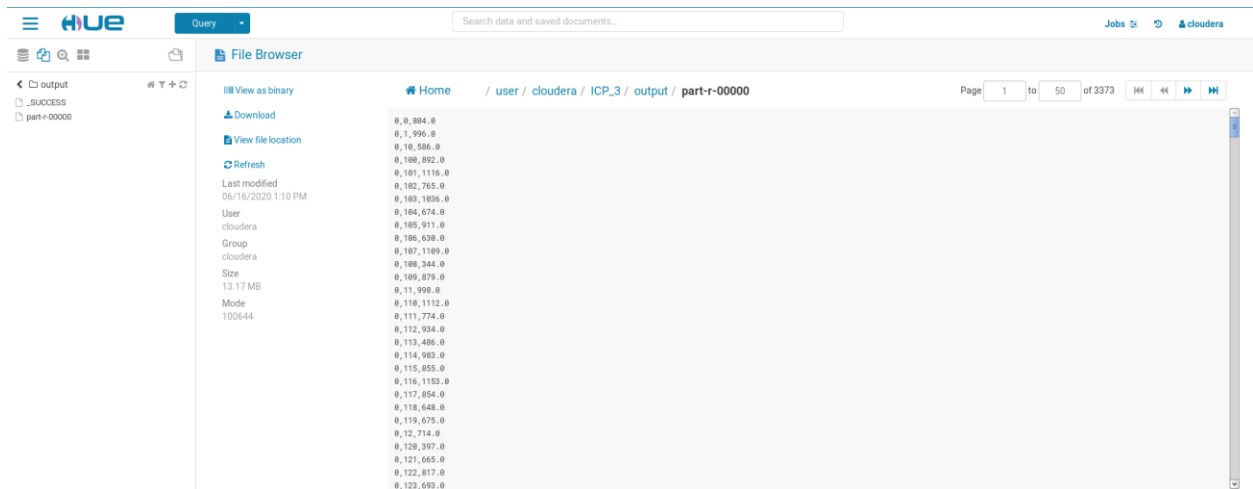
1 for each key  $(i, k)$  do
2   | sort values begin with  $M$  by  $j$  in  $list_M$ 
3   | sort values begin with  $N$  by  $j$  in  $list_N$ 
4   | multiply  $m_{ij}$  and  $n_{jk}$  for  $j_{th}$  value of each list
5   | sum up  $m_{ij} * n_{jk}$ 
6 return  $(i, k), \sum_{j=1} m_{ij} * n_{jk}$ 

```

Map-Reduce Job

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop jar MatrixMul.jar MatrixMul /user/cloudera/ICP_3  
/user/cloudera/ICP_3/output  
20/06/16 05:54:16 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0  
:8032  
20/06/16 05:54:16 WARN mapreduce.JobResourceUploader: Hadoop command-line option  
parsing not performed. Implement the Tool interface and execute your applicatio  
n with ToolRunner to remedy this.  
20/06/16 05:54:17 INFO input.FileInputFormat: Total input paths to process : 4  
20/06/16 05:54:17 WARN hdfs.DFSClient: Caught exception  
java.lang.InterruptedException  
    at java.lang.Object.wait(Native Method)  
    at java.lang.Thread.join(Thread.java:1281)  
    at java.lang.Thread.join(Thread.java:1355)  
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DF  
SOutputStream.java:967)  
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutpu  
tStream.java:705)  
    at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStre  
am.java:894)  
20/06/16 05:54:17 INFO mapreduce.JobSubmitter: number of splits:4  
20/06/16 05:54:17 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_15  
92303723860_0002  
20/06/16 05:54:18 INFO impl.YarnClientImpl: Submitted application application_15  
92303723860_0002  
20/06/16 05:54:18 INFO mapreduce.Job: The url to track the job: http://quickstar  
t.cloudera:8088/proxy/application_1592303723860_0002/  
20/06/16 05:54:18 INFO mapreduce.Job: Running job: job_1592303723860_0002  
20/06/16 05:54:31 INFO mapreduce.Job: Job job_1592303723860_0002 running in uber  
mode : false  
20/06/16 05:54:31 INFO mapreduce.Job: map 0% reduce 0%  
20/06/16 05:55:16 INFO mapreduce.Job: map 1% reduce 0%  
20/06/16 05:55:18 INFO mapreduce.Job: map 2% reduce 0%  
20/06/16 05:55:24 INFO mapreduce.Job: map 3% reduce 0%  
20/06/16 05:55:26 INFO mapreduce.Job: map 5% reduce 0%  
20/06/16 05:55:30 INFO mapreduce.Job: map 6% reduce 0%  
20/06/16 05:55:33 INFO mapreduce.Job: map 7% reduce 0%  
20/06/16 05:55:39 INFO mapreduce.Job: map 8% reduce 0%  
20/06/16 05:56:01 INFO mapreduce.Job: map 9% reduce 0%  
20/06/16 05:56:05 INFO mapreduce.Job: map 11% reduce 0%  
20/06/16 05:56:22 INFO mapreduce.Job: map 14% reduce 0%  
20/06/16 05:56:36 INFO mapreduce.Job: map 15% reduce 0%  
20/06/16 05:56:42 INFO mapreduce.Job: map 16% reduce 0%  
20/06/16 05:56:43 INFO mapreduce.Job: map 17% reduce 0%  
20/06/16 05:56:48 INFO mapreduce.Job: map 18% reduce 0%  
20/06/16 05:57:05 INFO mapreduce.Job: map 21% reduce 0%  
20/06/16 05:57:25 INFO mapreduce.Job: map 22% reduce 0%  
20/06/16 05:57:27 INFO mapreduce.Job: map 23% reduce 0%  
20/06/16 05:57:31 INFO mapreduce.Job: map 24% reduce 0%  
20/06/16 05:57:55 INFO mapreduce.Job: map 28% reduce 0%
```

Result:



The screenshot shows the Hive File Browser interface. The top navigation bar includes the Hive logo, a 'Query' dropdown, a search bar, and a 'Jobs' link. The left sidebar shows a directory tree with 'output' expanded, containing '_SUCCESS' and 'part-r-00000'. The main area displays details for 'part-r-00000' at the path '/ user / cloudera / ICP_3 / output / part-r-00000'. The file is 13.17 MB, last modified on 06/16/2020 at 1:10 PM, and is owned by 'cloudera'. The file content is displayed as a large grid of floating-point numbers.

File Name	Size	Last Modified	User	Group
part-r-00000	13.17 MB	06/16/2020 1:10 PM	cloudera	cloudera

0.0, 884.0	0.1, 996.0	0.10, 586.0	0.100, 892.0	0.101, 116.0	0.102, 765.0	0.103, 1036.0	0.104, 674.0	0.105, 011.0	0.106, 638.0	0.107, 1199.0	0.108, 344.0	0.109, 879.0	0.11, 998.0	0.110, 1112.0	0.111, 774.0	0.112, 934.0	0.113, 486.0	0.114, 983.0	0.115, 855.0	0.116, 1153.0	0.117, 854.0	0.118, 648.0	0.119, 675.0	0.12, 714.0	0.120, 397.0	0.121, 665.0	0.122, 817.0	0.123, 693.0
------------	------------	-------------	--------------	--------------	--------------	---------------	--------------	--------------	--------------	---------------	--------------	--------------	-------------	---------------	--------------	--------------	--------------	--------------	--------------	---------------	--------------	--------------	--------------	-------------	--------------	--------------	--------------	--------------

Reference:

<https://umkc.app.box.com/s/xk4jj0do3p7fa3swx6utcuazx3wny8j8>

<https://lendap.wordpress.com/2015/02/16/matrix-multiplication-with-mapreduce/>