ICP3 - Hadoop and Map Reduce

CS5590 - Summer 2020

MATRIX MULTIPLICATION

Map Function

Algorithm 1: The Map Function

- 1 for each element mij of M do
- produce (key, value) pairs as $((i, k), (M, j, m_{ij}))$, for k = 1, 2, 3, ... up to the number of columns of N
- 3 for each element njk of N do
- 4 produce (key, value) pairs as $((i, k), (N, j, n_{jk}))$, for i = 1, 2, 3, ... 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

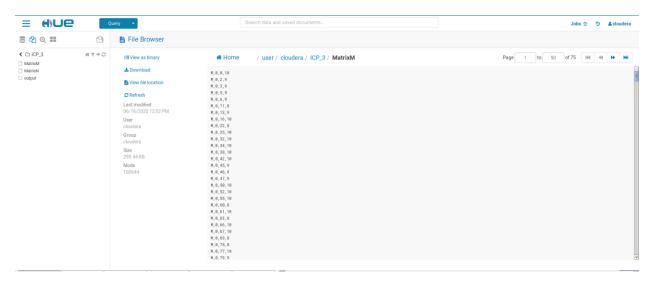
- 1 for each key (i,k) do
- sort values begin with M by j in $list_M$
- sort values begin with N by j in $list_N$
- 4 multiply m_{ij} and n_{jk} for j_{th} value of each list
- sum up $m_{ij} * n_{jk}$
- 6 return $(i,k), \sum\limits_{j=1} m_{ij}*n_{jk}$

Hadoop jar MatrixMul.jar MatrixMul /user/cloudera/ICP 3 /user/cloudera/ICP 3/output

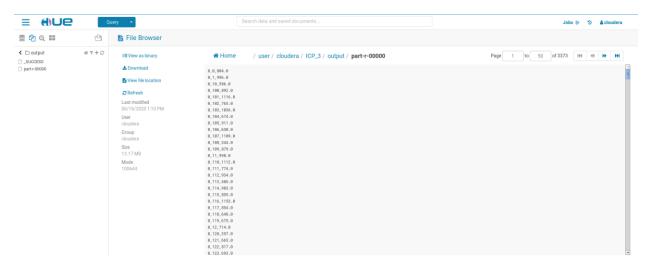
```
cloudera@quickstart:~
                                                                                   _ 🗆 ×
File Edit View Search Terminal Help
[cloudera@quickstart ~] $ hadoop jar MatrixMul.jar MatrixMul /user/cloudera/ICP 3
/user/cloudera/ICP 3/output
20/06/16 05:54:16 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0
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%
                                      map 17% reduce 0%
20/06/16 05:56:43 INFO mapreduce.Job:
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%
```

SCREENSHOTS

Cloudera:



Result:



Reference:

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

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