

ELEN4020A: Data Intensive Computing Lab 3

Jared Gautier (820687), Nick Raal (793528), Sasha Berkowitz (818737), Arunima Pathania()

I. INTRODUCTION

The objective of the lab was to perform matrix multiplication using various frameworks and MapReduce. The MapReduce framework was introduced by Google to support distributed computing on large data sets onto clusters of computers. The data is replicated multiple times in parallel on the system for increased efficiency, reliability and availability. This was done by using the MrJob framework adapted to two different algorithms in the the Python language.

II. ALGORITHMS

The reason for using the MapReduce framework is due to the benefits it provides. MapReduce has the ability to take a query over a data set, divide it, then run the query in parallel over multiple nodes. This benedits in removing the issue of small computers processing data too large to handle, using multiple servers and the Batch processing model.

A. Map

The main purpose of the map function is to generate a key,value pair. The map function takes individual tasks and transforms the input records into intermediate records, which can be processed by the multiplication algorithm. Each of these transformations occur in parallel. The map function shuffles the key, value pairs based on the first key to re-organize the output.

B. Reduce

The reduce function takes the re-organized data from the map function and reduces them to a summarized data set, the desired output. The reduce function, in terms of this laboratory, performs matrix multiplication using the generated key,value pair.This is done by multiplying the values with the keys and storing them. The final value is the sum of the different products obtained in the previous step.

C. Algorithm A

- 1) *Mapper*: The mapper algorithm that was used
- 2) *Reducer*:

III. RESULTS

Algorithm 1 The mapper function

Require: Map function to produce and return key, value pairs

```

for  $value\_A \leftarrow 0$  to  $A$  do
   $k \leftarrow 1$  to  $B$ 
   $((i, k), (A, j, value\_A))$  for each value of  $k$ 
end for
for  $value\_B \leftarrow 0$  to  $B$  do
   $i \leftarrow 1$  to  $A$ 
   $((i, k), (B, j, value\_B))$  for each value of  $i$ 
end for
return  $(key, value)$  pair

```

Algorithm 2 The Reducer Function

Require: Uses output of Mapper function to perform matrix multiplication

```

for  $(i, k) \leftarrow 0$  to  $A$  do
  sort  $(A, j, value\_A)$  by  $j$ 
  sort  $(B, j, value\_B)$  by  $j$ 
   $value\_A * j \leftarrow multiA$ 
   $value\_B * j \leftarrow multiB$ 
   $Answer \leftarrow multiA + multiB$ 
  Return  $(i, k), Answer$ 
end for

```
