Distributed Data Analytics

Lab 1

Submitted By- Abdullah Al Foysal

Id:277458

Exercise 0: Explain the system

Processor	Intel(R)Core(TM)i5-8250U CPU@ 1.60 GHz 1.80 GHz
Number of cores	4
RAM	8 GB
OS	Windows
Programming Language Version	Python 3.5

Exercise 1: Basic Parallel Vector Operation with MPI

a) Adding two vectors and storing in third vector

Here, we have initialised two random vector with sizes={10^7}.

We split the vector by np.array_split(). Process 0 will be the master and split and distribute the vector.Then, we have the split vectors and distribute it into the processes from (1 to N-1). Process 0 will also do a part of the vector addition by taking the first partition and adding it.

Here, we send the splitted vector to rest of the processes. At the same time, we will be receiving the data from the rest of the processes.

The child nodes will receive chunks of vectors and does the addition. Then, it sends back to process 0. Process 0 will receive the chunks and club it into one vector.

10^7 with 2 processors	20.36
10^7 with 3 processors	27.51
10^7 with 8 processors	35.89
10^12 with 2 processors	Memory error
10^12 with 3 processors	Memory error
10^12 with 8 processors	Memory error
10^15 with 3 processors	Memory error
10^15 with 8 processors	Memory error

b) Average of the vector

We have to the same as above. We have to split the vector into chunks. Process 0 will do the first task, mean. The child nodes does the mean and give it to the process 0 for concatenating the mean of the chunks of vector and then averaging the mean of all the means of vector chunks.

10^7 with 2 processors	0.268
10^7 with 3 processors	0.257
10^7 with 8 processors	0.34
10^12 with 2 processors	Memory error
10^12 with 3 processors	Memory error
10^12 with 8 processors	Memory error
10^15 with 3 processors	Memory error
10^15 with 8 processors	Memory error

Exercise 2: Parallel vector multiplication using MPI

Here, we need to split the vector but row wise and then do the dot product of row vector of vector 1 and the column of vector 2. This will be done by splitting the row and column vector and then distributing the chunks for dot product. Then, we have received the data and add them to find one vector value. Then, we append the value to new column vector which will be our final vector.

10^2 with 3 processors	0.0208
10^2 with 8 processors	0.0686
10^3 with 3 processors	0.0207
10^3 with 8 processors	0.6961
10^4 with 3 processors	3.9477
10^4 with 8 processors	11.43

Exercise 3: Parallel Matrix Multiplication using MPI

Here, we need to use scatter and gather function. Here, also we should split row and column vector by sending it to all processes. This happens in process 0, it should scatter and also do the operation of dot product of the vectors. Then, it should gather the dot product of all the other processes and add it to the get one element A[row][column].

10^2 with 3 processors	1.5917
10^2 with 8 processors	7.1476
10^3 with 3 processors	243.47
10^3 with 8 processors	It was taking too much time to run the code
10^4 with 3 processors	It was taking too much time to run the code
10^4 with 8 processors	It was taking too much time to run the code