CEG-7380

Cloud Computing

Spring 2016

Project #2

Report File

DHRUVKUMAR NAVINCHANDRA PATEL

U00791652

Matrix Multiplication in MapReduce

Following are the different steps which I followed during development of this project

1) **Analysis:** Here the idea is to perform a map Reduce Algorithm on two matrices to find a matrix multiplication and produce a desired output. For example here one Matrix named A contains I number of rows and j number of columns and other Matrix named B contains j number of rows and k number of columns. Resulting Matrix contains I number of rows and k number of columns. Main objective is to implement matrix multiplication with ONE map-reduce step.

1) **Hadoop MapReduce implementation :**

i) In Hadoop MapReduce I am taking sample_input.csv file as an input. I developed this program in our virtual Cloud which has a version of Hadoop 1.2.1. Also I created one sample java program to generate input file with any size of two matrices. I checked my output with small input and it's correct and then I checked for one matrix with (10×10) and second matrix with (10×5). Also the matrix value does not contains 0. All the values are real.

Input Format Example:

[matrix_name][,][row_index][,][column_index][,][value]

A,0,1,12.0

B,0,1,14.0

ii) **Mapper Implementation:** In Mapper I am taking data from input file and also take matrices size from configuration which is set in Driver class. Generate key and value pair in mapper and pass to reducer.

a) In mapper generate key and value as Text.

b) Here I split the data and differentiate according to Matrix A and Matrix B.

c) So Output from mapper is <I,k> as key for both matrices and A,j,aij value from first matrix and B,j,bjk from second matrix. Where i×j is the size of Matrix A and j×k is the size of Matrix B.

iii) **Reducer implementation:** In reducer I am taking the key and value from mapper and generate output in key as Text and value as DoubleWritable.

a) In reducer I have size of all the matrices and values according to key. So, I iterate through values and store into two dimensional array according to position so, I don't need to sort the data according to j.

b) Simple calculation of matrix multiplication from first row of one matrix from matrix 1 and first column of second matrix from matrix 2.

Store result as value in double in output file. Key as it is which emitted from mapper.

c) In Hadoop 1.2.1 version there is no TextOutputFormat.SEPERATOR so in my output there is a space between key and value.

Output from Virtual Cloud:

[row_index][,][column_index][,][value]

0,0 100.0

d) In Hadoop 2.6.0 version there is TextOutputFormat.SEPERATOR so I removed space between key and value and I am getting output exactly what you want.

Output is following format:

[row_index][,][column_index][,][value]

0,0,100.0

iv) **Driver class –** It contains all the required parameter to execute Map-Reduce Program. Here we have to set explicitly in Configuration parameter about matrix size.