

Matrix Vector Multiplication

Problem

Calculate the product of a matrix M (assumed sparse) and a vector v

Here we show a toy example

$$\begin{pmatrix} 0 & 2.5 & 1 & 0 \\ 4 & 0 & 0 & -2 \\ 0 & 0 & -1.5 & 0 \\ 8 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \\ -3 \\ 8 \end{pmatrix}$$

Workflow - 2 MapReduce jobs

1. CellMultiplication: Mapper1 + Mapper2 ---> Reducer
 - Mapper1: read the non-zero elements in the matrix file (row, col, $M[\text{row}][\text{col}]$)
input: <offset, line>
output: <col, row= $M[\text{row}][\text{col}]$ >
 - Mapper2: read the vector
input: <offset, line>
output: <row, $v[\text{row}]$ >
 - Reducer: multiply a matrix column with the corresponding vector row
input: <col, (row1= $M[\text{row1}][\text{col}]$, row2= $M[\text{row2}][\text{col}]$, ..., $v[\text{col}]$)>
output: <row, $M[\text{row}][\text{col}] * v[\text{col}]$ >
2. CellSum: Mapper ---> Reducer
 - Mapper: read the intermediate result of cell multiplication
input: <offset, line>
output: <row, $M[\text{row}][\text{col}] * v[\text{col}]$ >
 - Reducer: sum up all the cell product to the final value for each vector row
input: <row, ($M[\text{row}][\text{col1}] * v[\text{col1}]$, $M[\text{row}][\text{col2}] * v[\text{col2}]$, ...)>
output: <row, $M[\text{row}][\text{col1}] * v[\text{col1}] + M[\text{row}][\text{col2}] * v[\text{col2}] + \dots$ >

Note

1. This example uses two MapReduce jobs, and the later takes the output of the former as the input. Note the dependency of input/output directories.
2. In the first MapReduce job, two Mapper classes are used to read multiple input data, which is a good demo for the usage of **MultipleInputs** class in package *org.apache.hadoop.mapreduce.lib.input*.

Input

Matrix.txt

1	2	2.5
1	3	1
2	1	4
2	4	-2
3	3	-1.5
4	1	8

Vector.txt

1
0
2
-1