2017 Introduction to Massive Data Analysis

Assignment 1

✧ **Deadline: 2017.3.31(Fri.) 23:59**

Please write a **MapReduce** program in Hadoop to solve the following question.

Question: Matrix Multiplication

If ***M*** is a matrix with element m_{ij} in row *i* and column *j*, and ***N*** is a matrix with element n_{jk} in row *j* and column *k*, then the product ***P*** = ***MN*** is the matrix ***P*** with element p_{ik} in row *i* and column *k*, where

$$p_{ik} = \sum_j m_{ij} n_{jk}$$

Data format

Input:

```
M,0,0,10
M,0,1,0
M,0,2,20
M,1,0,0
M,1,1,30
M,1,2,0
M,2,0,40
M,2,1,0
M,2,2,50
N,0,0,1
N,0,1,2
N,0,2,3
N,1,0,4
N,1,1,5
N,1,2,0
N,2,0,6
N,2,1,7
N,2,2,8
```

M(i × j)

10	0	20
0	30	0
40	0	50

N (j × k)

1	2	3
4	5	0
6	7	8

i,j,k $\leqq$ 1000

$0 \leqq m_{ij}, n_{jk} \leqq 1000$

Output:

130	160	190
120	150	0
340	430	520

The output data set containing these matrices are represented as follows.

```
0,0,130
0,1,160
0,2,190
1,0,120
1,1,150
1,2,0
2,0,340
2,1,430
2,2,520
```

Assignment Requirements:

1. Code(.java)
2. Output data(.txt) ┐ 80%

Note: Output data is calculated from **500input.txt(two 500×500 matrices)** which we provided.

3. Report(.pdf) ---------- 20% : Explain how you design your mapper and reducer.

Please pack the above files into a zip file. Name it as "MDA_HW1_studentID.zip".