

語言

python3

環境

pyspark

執行

Python3 HW1.py

方法

Map:

將矩陣 M N 透過 map function 成 key-value pair $(j, (i, M_{ij}))$ 和 $(j, (k, N_{jk}))$

Reduce:

將相同 key 的所有 value 相加

步驟：

step 1 將 M N 矩陣 透過 map function 轉成 key-value pairs
 $M : M_{ij} \Rightarrow (j, (i, M_{ij}))$
 $N : N_{jk} \Rightarrow (j, (k, N_{jk}))$

step 2 Join M 跟 N 的 key-value pair :
 $(j, (i, M_{ji})) \text{ join } (j, (k, N_{jk})) \Rightarrow (j, ((i, M_{ij}), (k, N_{jk})))$

step 3 將 step2 後的 key-value pairs 透過 map function 成 $((i, k), v)$
 其中 $v = M_{ij} * N_{jk}$

step 4 Reduce the key-value pairs, 將相同的 key 的 v 總和

step 5 針對 key 排列所有的 pairs (row to column)

step 6 輸出成 output.txt