

MDA Assignment 1 Report

我在這次作業使用了三個mapper及一個reducer

mapper1：這個mapper用來處理input data，由於M陣列的第一個數字和N陣列的第二個數字是用來配對M、N且之後用不到（ $M_{ij} * N_{jk} = M_{Nik}$ ），故將先取出且放在mapper key的位置。value的位置則是用一個list來放 $[M, i, M_{ij}]$ 或 $[N, k, N_{jk}]$ ，由於M、N的狀況不同，用if-else處理。

reducer1：reducer1則是將兩個傳進來的值相加，這題恰巧只需要一個reducer，一次用來作list相加，一次用在值相加。

mapper2：這個mapper主要做演算法的部分，前面的key為一個tuple，內容為(i,k)，value的部分則是 $M_{ij} * N_{jk}$ ，由於前面mapper1及reducer1做完後的value是前面有500個M項，後面有500個N項，所以我用兩層for-loop從前面從第一項開始(第一個M的i)，配對後面從第1501項（為N的第一個k）開始做append。

mapper3：這個mapper在做檔案寫入的動作，由於fwrite內只能寫入string，所以將line[0][0]、line[0][1]及line[1]轉成string後相加，也加入逗號，就完成最終的OutputFile了。

整題的流程為先使用mapper1後使用reduceByKey(reducer1)將後面的list相加，lines會變成（j, $[M, i, M_{ij}, \dots, N, k, N_{jk}, \dots]$ ），再使用mapper2，做矩陣內對應位置值相乘的動作，再用reducer1將值相加，最終寫入OutputFile即可完成這次作業。