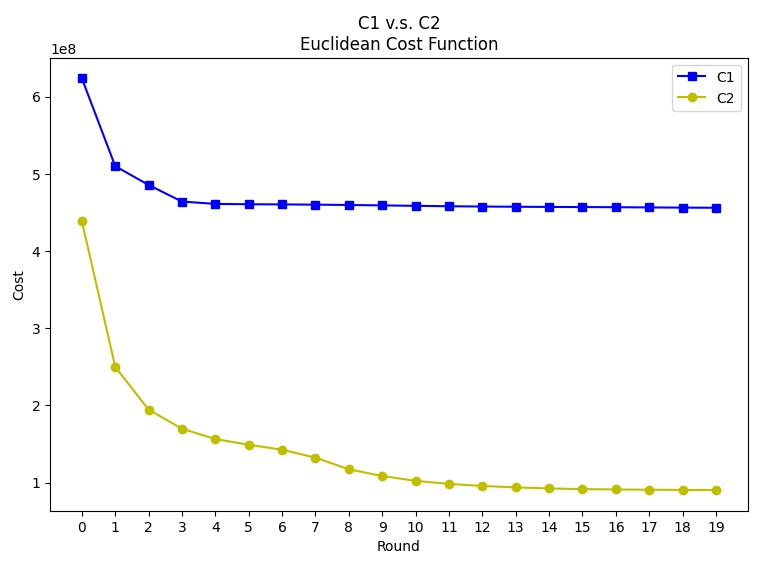
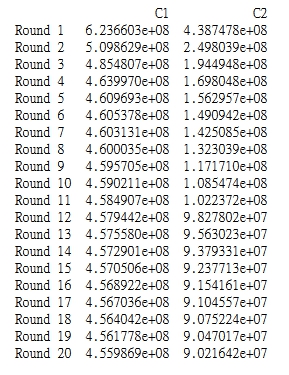
**k-means on MapReduce**

105072123 黃海茵

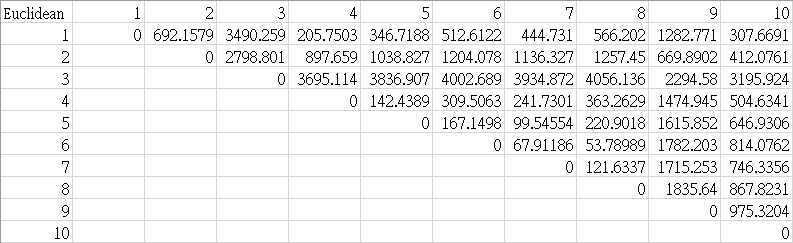
1. Euclidean



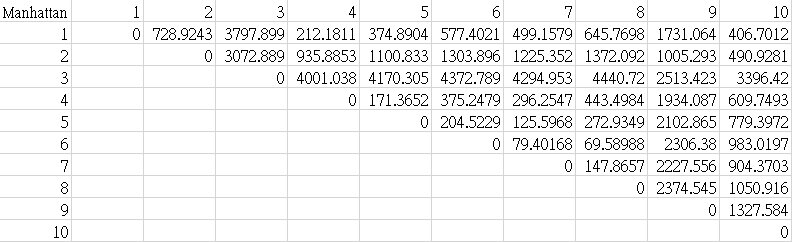
* Percentage improvement：C1是26.885%，C2是79.438%。

C2各點相距較遠，所以data在最初的距離會較C1來的大，因此C2的percentage improvement會較大。使用Euclidean時，是計算平方和開根號，也就是兩點間的距離，所以centroid相距較遠，cluster分布會較平均，每個點到centroid的距離也會越小，所以C2的cost會小於C1。

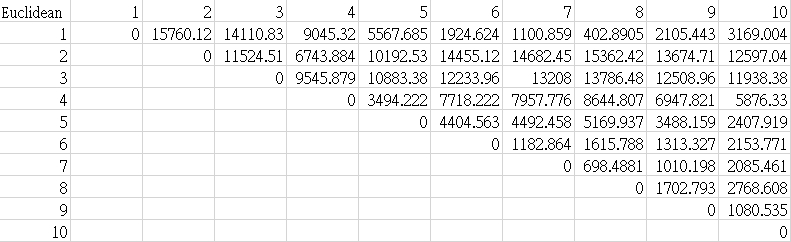
* C1 Euclidean



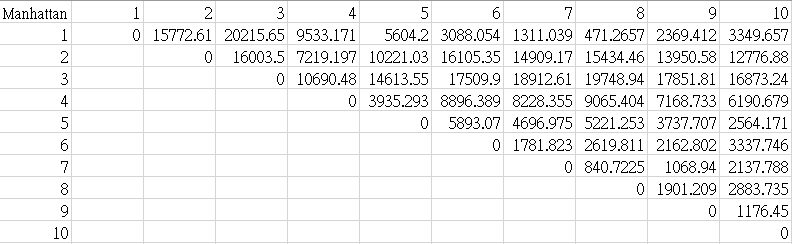
* C1 Manhattan



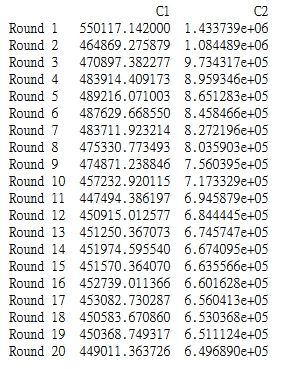
* C2 Euclidean

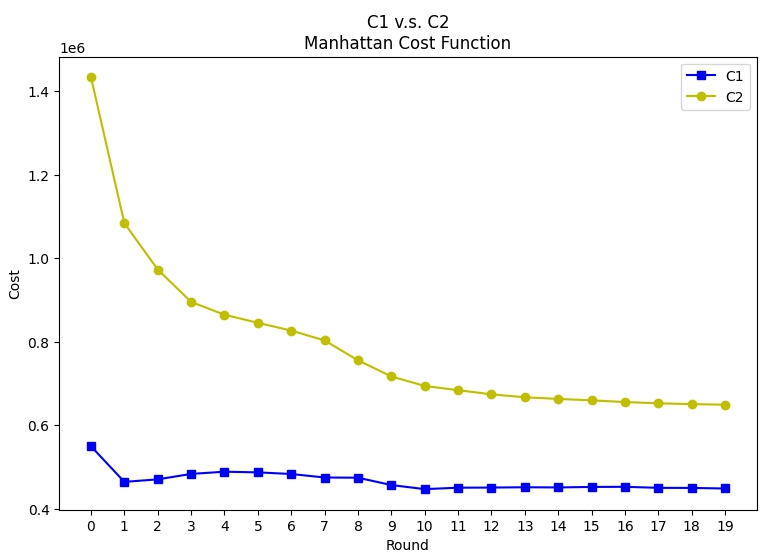


* C2 Manhattan



1. Manhattan

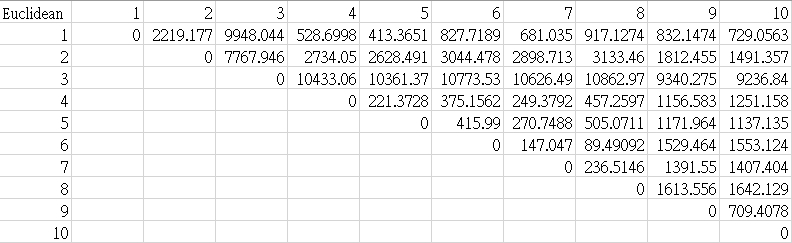




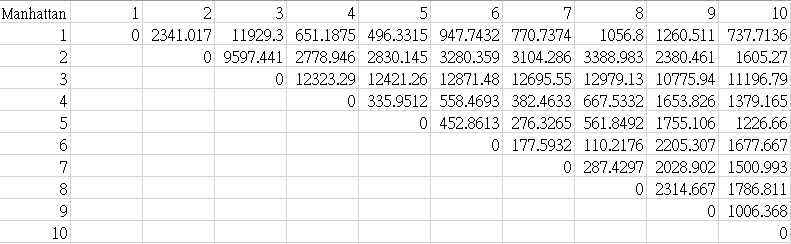
* Percentage improvement：C1是18.379%，C2是54.686%。

同(a)小題，C2各點相距較遠，所以data在最初的距離會較C1來的大，因此C2的percentage improvement會較大。使用Manhattan時，是計算每個維度距離的總和，而非兩點間的距離，所以不一定會比較小。

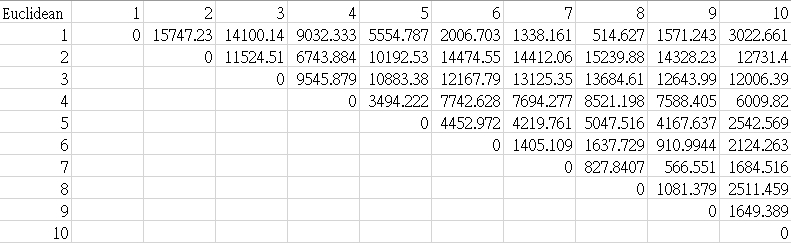
* C1 Euclidean



* C1 Manhattan



* C2 Euclidean



* C2 Manhattan

