Report
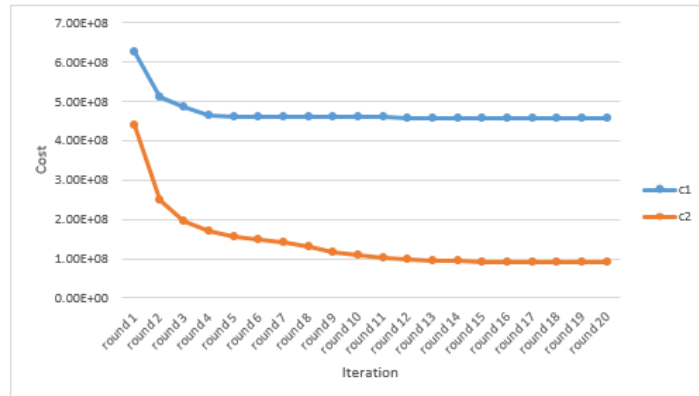
指令：yarn jar kmeansjava-1.0-SNAPSHOT.jar org.apache.hadoop.examples.Kmeans c1.txt c2.txt

1. (a) 和 (b) 的 cost vs. iteration 圖
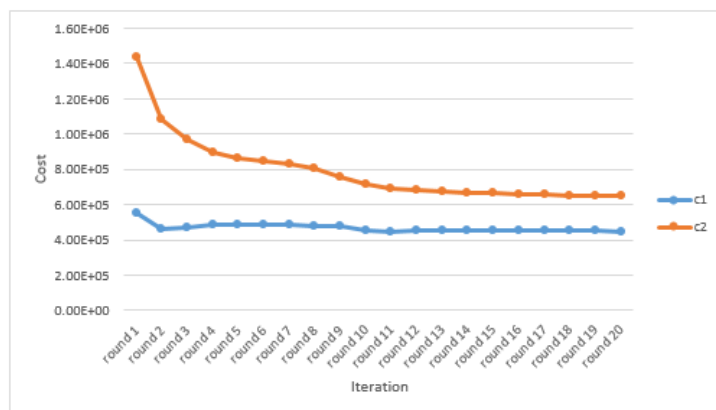
Euclidean

	c1	c2
round 1	6.24E+08	4.39E+08
round 2	5.10E+08	2.50E+08
round 3	4.85E+08	1.94E+08
round 4	4.64E+08	1.70E+08
round 5	4.61E+08	1.56E+08
round 6	4.61E+08	1.49E+08
round 7	4.60E+08	1.43E+08
round 8	4.60E+08	1.32E+08
round 9	4.60E+08	1.17E+08
round 10	4.59E+08	1.09E+08
round 11	4.58E+08	1.02E+08
round 12	4.58E+08	9.83E+07
round 13	4.58E+08	9.56E+07
round 14	4.57E+08	9.38E+07
round 15	4.57E+08	9.24E+07
round 16	4.57E+08	9.15E+07
round 17	4.57E+08	9.10E+07
round 18	4.56E+08	9.08E+07
round 19	4.56E+08	9.05E+07
round 20	4.56E+08	9.02E+07

Manhattan

	c1	c2
round 1	5.50E+05	1.43E+06
round 2	4.65E+05	1.08E+06
round 3	4.71E+05	9.73E+05
round 4	4.84E+05	8.96E+05
round 5	4.89E+05	8.65E+05
round 6	4.88E+05	8.46E+05
round 7	4.84E+05	8.27E+05
round 8	4.75E+05	8.04E+05
round 9	4.75E+05	7.56E+05
round 10	4.57E+05	7.17E+05
round 11	4.47E+05	6.95E+05
round 12	4.51E+05	6.84E+05
round 13	4.51E+05	6.75E+05
round 14	4.52E+05	6.67E+05
round 15	4.52E+05	6.64E+05
round 16	4.53E+05	6.60E+05
round 17	4.53E+05	6.56E+05
round 18	4.51E+05	6.53E+05
round 19	4.50E+05	6.51E+05
round 20	4.49E+05	6.50E+05

2. 做完 10 次 iteration 之後 (a) 和 (b) 的 Percentage improvement values

Euclidean：

C1：

(623660345.306413 - 459021103.342289) / 623660345.306413 =

0.263988632920431

C2：

(438747790.027914 - 108547377.178569) / 438747790.027914 =

0.752597324372476

　　從 cost 改善的情況可以發現，C2 比 C1 有更明顯的改善；而從 cost 來看，C2 一直都比 C1 好。所以，C2 是比較好的 initialization。

Manhattan：

C1：

(550117.142000001 - 457232.920115077) / 550117.142000001 =

0.168844442016904

C2：

(1433739.31 - 717332.902543228) / 1433739.31 = 0.499676895555561

　　從 cost 改善的情況可以發現，C2 比 C1 有更明顯的改善；但若從 cost 來看，C1 一直都比 C2 好。所以，若以 cost 改善的觀點來看，C2 是比較好的 initialization；而從 cost 來看，C1 比較好。

3. 做完 20 次 iterations 之後，所有 Cluster Centroids 兩兩之間的 Euclidean 和 Manhattan Distance

C1

Euclidean-Euclidean distance

```
0-1 692.1578865536187
0-2 3490.2586403239793
0-3 205.75827883457204
0-4 346.71882253414816
0-5 512.6122467083617
0-6 444.7310005470825
0-7 566.2019922936282
0-8 1282.7708445154385
0-9 307.66912835221507
1-2 2798.8010531589434
1-3 897.6589863450828
1-4 1038.8268802911389
1-5 1204.0781989863117
1-6 1136.3273438149283
1-7 1257.4495275592424
1-8 669.8902282318555
1-9 412.0760077167744
2-3 3695.114191079636
2-4 3836.9066381524567
2-5 4002.6890825730284
2-6 3934.8715588311575
2-7 4056.1355729719103
2-8 2294.57964158953
2-9 3195.9239010108936
3-4 142.43887392408647
3-5 309.50632446700047
3-6 241.7301145044183
3-7 363.2628951046584
3-8 1474.9454213642691
3-9 504.63411599571515
4-5 167.1498001315246
4-6 99.54554331498562
4-7 220.90178372040566
4-8 1615.852353440387
4-9 646.9305638786212
5-6 67.91186107588504
5-7 53.78989116172683
5-8 1782.2030486002302
5-9 814.0761501339795
6-7 121.63372043718282
6-8 1715.2531997144126
6-9 746.3355586141014
7-8 1835.6396718448902
7-9 867.8230790917461
8-9 975.3204225971708
```

Euclidean-Manhattan distance

```
0-1  728.9243139995409
0-2  3797.8990780180675
0-3  212.18109038688195
0-4  374.8904224625678
0-5  577.4020758222025
0-6  499.15789394709685
0-7  645.7697774866793
0-8  1731.0643066429416
0-9  406.70122510554586
1-2  3072.8886904466503
1-3  935.8853379058191
1-4  1100.833091013825
1-5  1303.8957233218873
1-6  1225.351713180189
1-7  1372.0922054892176
1-8  1005.2930456989247
1-9  490.928058154678S
2-3  4001.0380519416276
2-4  4170.304532612548
2-5  4372.788718801669
2-6  4294.952834221333
2-7  4440.719767637059
2-8  2513.422660256411
2-9  3396.4200031055902
3-4  171.36515415561516
3-5  375.2479208943975
3-6  296.25472354226696
3-7  443.49844537759145
3-8  1934.086959807074
3-9  609.749321623295
4-5  204.5229364239253
4-6  125.59678617977005
4-7  272.93491284339433
4-8  2102.8649228110594
4-9  779.3972267080745
5-6  79.40168444202222
5-7  69.58987631971816
5-8  2306.3802505970652
5-9  983.019680661424
6-7  147.86570909768562
6-8  2227.555856782107
6-9  904.3702500156849
7-8  2374.545430478822
7-9  1050.9162214834769
8-9  1327.5039795031056
```

Manhattan-Euclidean distance

```
0-1 2219.1772770509865
0-2 9948.044077639335
0-3 528.6597575475606
0-4 413.3650612071161
0-5 827.7188856579816
0-6 681.0349895443137
0-7 917.127382961168
0-8 832.1474343052389
0-9 729.0563485517976
1-2 7767.945602579569
1-3 2734.0498544640413
1-4 2628.4908097297634
1-5 3044.4778721252733
1-6 2898.7128939326617
1-7 3133.460130173686
1-8 1812.4545744763466
1-9 1491.3573457511582
2-3 10433.061351319797
2-4 10361.367486044566
2-5 10773.530838116945
2-6 10626.48859681842
2-7 10862.965776276442
2-8 9340.275232170257
2-9 9236.840021697923
3-4 221.37279398782263
3-5 375.1561884837036
3-6 249.3791882654921
3-7 457.25965255434755
3-8 1156.5833757939854
3-9 1251.1583460671166
4-5 415.9899852468605
4-6 270.74879157775985
4-7 505.07106661337866
4-8 1171.964285896139
4-9 1137.135265913954
5-6 147.04697388149492
5-7 89.4989165524104
5-8 1529.4640115204897
5-9 1553.1238066646654
6-7 236.51462239245419
6-8 1391.550421146714
6-9 1407.404400267174
7-8 1613.555789411286
7-9 1642.1286873773308
8-9 709.4077855501387
```

Manhattan-Manhattan distance

```
0-1 2341.017218605991
0-2 11929.300151785714
0-3 651.187488095238
0-4 496.33152119047617
0-5 947.7432364658634
0-6 770.7373834915885
0-7 1056.7995806298163
0-8 1260.5105607142857
0-9 737.7135730519481
1-2 9597.441187096774
1-3 2778.9457620967737
1-4 2830.1445281720426
1-5 3280.3591681796074
1-6 3104.2857711482065
1-7 3388.9826482565923
1-8 2380.460958064516
1-9 1605.2701287064187
2-3 12323.287569444445
2-4 12421.263079999999
2-5 12871.483428978225
2-6 12695.554202097901
2-7 12979.13318018672
2-8 10775.939185714284
2-9 11196.78698181818
3-4 335.951213
3-5 558.4692581658292
3-6 382.4633301282051
3-7 667.5332295988935
3-8 1653.8258869047618
3-9 1379.165172979798
4-5 452.86133068676713
4-6 276.3264914965834
4-7 561.8492485408023
4-8 1755.105532761905
4-9 1226.6603546666668
5-6 177.59316237363976
5-7 110.21762404686818
5-8 2205.3073830102894
5-9 1677.666863534846
6-7 287.42970773365636
6-8 2028.901615784216
6-9 1500.9934102564105
7-8 2314.667454756965
7-9 1786.8113163428895
8-9 1006.3678264069263
```

C2

Euclidean-Euclidean distance

```
0-1 15760.12247224593
0-2 14110.83439088334
0-3 9045.32023455239
0-4 5567.684524118405
0-5 1924.6248815733204
0-6 1100.8590503593277
0-7 402.89054961102914
0-8 2105.4425755929597
0-9 3169.003772849924
1-2 11524.505650179837
1-3 6743.88410019246
1-4 10192.525007384162
1-5 14455.11937212132
1-6 14682.458992891092
1-7 15362.417968805125
1-8 13674.707531226235
1-9 12597.039559531104
2-3 9545.879403387185
2-4 10883.382187801386
2-5 12233.959804503613
2-6 13208.002933714424
2-7 13786.484182516106
2-8 12508.95738096874
2-9 11938.376127029482
3-4 3494.2224155718104
3-5 7718.222009696623
3-6 7957.775949135465
3-7 8644.807041005524
3-8 6947.820636329124
3-9 5876.330199605868
4-5 4404.562590797219
4-6 4492.458214368271
4-7 5169.9372911256405
4-8 3488.1585187816186
4-9 2407.918794485802
5-6 1182.864189045907
5-7 1615.7882361392917
5-8 1313.3274934048693
5-9 2153.771471752397
6-7 698.4881359277906
6-8 1010.1976652126849
6-9 2085.4606764072964
7-8 1702.7926583344838
7-9 2768.6077191659247
8-9 1080.534943935508
```

Euclidean-Manhattan distance

```
0-1  15772.614899885255
0-2  20215.645980206544
0-3  9533.17084939759
0-4  5604.200489099255
0-5  3088.0543184882386
0-6  1311.039156985083
0-7  471.26571999995156
0-8  2369.4121590413606
0-9  3349.657086008631
1-2  16003.499
1-3  7219.196666666668
1-4  10221.031
1-5  16105.3475
1-6  14909.169518714286
1-7  15434.460040787624
1-8  13950.575945454546
1-9  12776.883065217391
2-3  10690.484333333334
2-4  14613.552
2-5  17509.90275
2-6  18912.605418714285
2-7  19748.93569338959
2-8  17851.806836363638
2-9  16873.243673913043
3-4  3935.2926666666663
3-5  8896.389208333332
3-6  8228.355075
3-7  9065.404333333334
3-8  7168.732963636363
3-9  6190.679311594203
4-5  5893.070125
4-6  4696.975382142857
4-7  5221.252805987173
4-8  3737.707
4-9  2564.1705434782607
5-6  1781.8226714285715
5-7  2619.8113862517584
5-8  2162.8021454545456
5-9  3337.7462608695655
6-7  840.7225236939923
6-8  1068.9399724025977
6-9  2137.7882566770186
7-8  1901.208756322721
7-9  2883.734536812817
8-9  1176.4504256916996
```

Manhattan-Euclidean distance

```
0-1 15747.234225995131
0-2 14100.144687804226
0-3 9832.333822974994
0-4 5554.786693382845
0-5 2006.7026683699517
0-6 1338.16112552521
0-7 514.627037539365
0-8 1571.2434198098995
0-9 3022.6608840606823
1-2 11524.505650179837
1-3 6743.88410019246
1-4 10192.525087384162
1-5 14474.554115568231
1-6 14412.056615494355
1-7 15239.87707116475
1-8 14328.226191823898
1-9 12731.397634823861
2-3 9545.879403387185
2-4 10883.382187801386
2-5 12167.79387138704
2-6 13125.351004065811
2-7 13684.606757319763
2-8 12643.985638343242
2-9 12006.39461776129
3-4 3494.2224155718104
3-5 7742.6281172713425
3-6 7694.276701483224
3-7 8521.197863130883
3-8 7588.404540222487
3-9 6009.820222838373
4-5 4452.971684507903
4-6 4219.760574098285
4-7 5847.516256062294
4-8 4167.636533016605
4-9 2542.5693542004556
5-6 1405.1090803330765
5-7 1637.7294382123869
5-8 910.9943878357352
5-9 2124.2633623517277
6-7 827.840658014979
6-8 566.5510174100833
6-9 1684.516011855046
7-8 1881.3793348240201
7-9 2511.458858671225
8-9 1649.3891719444812
```

Manhattan-Manhattan distance

```
[root@sandbox target]# hea
0-1  15757.691264458323
0-2  20200.259435786043
0-3  9517.668232509728
0-4  5588.853634537063
0-5  3281.488246764413
0-6  1430.2086778758846
0-7  602.9548488263027
0-8  2102.553977579326
0-9  3211.4557560606645
1-2  16083.499
1-3  7219.196666666668
1-4  10221.031
1-5  16325.2705
1-6  14506.48588961039
1-7  15335.957402597402
1-8  14980.056095890412
1-9  12922.931357142857
2-3  10690.484333333334
2-4  14613.552
2-5  17521.517666666663
2-6  18775.12146103896
2-7  19602.262814935064
2-8  18111.885424657536
2-9  16995.133535714285
3-4  3935.2926666666663
3-5  9116.0245
3-6  8090.510188311688
3-7  8918.813116883115
3-8  7771.22207762557
3-9  6312.5300119047615
4-5  6110.8325
4-6  4293.5019025974025
4-7  5123.066808441558
4-8  4768.923
4-9  2710.0564999999997
5-6  1855.5799090909088
5-7  2682.5692337662335
5-8  1358.795894977169
5-9  3413.0361785714285
6-7  833.4302824675323
6-8  674.8275699163852
6-9  1784.5120454545456
7-8  1500.824883695072
7-9  2613.9973051948054
8-9  2062.251068003914
[root@sandbox target]#
```