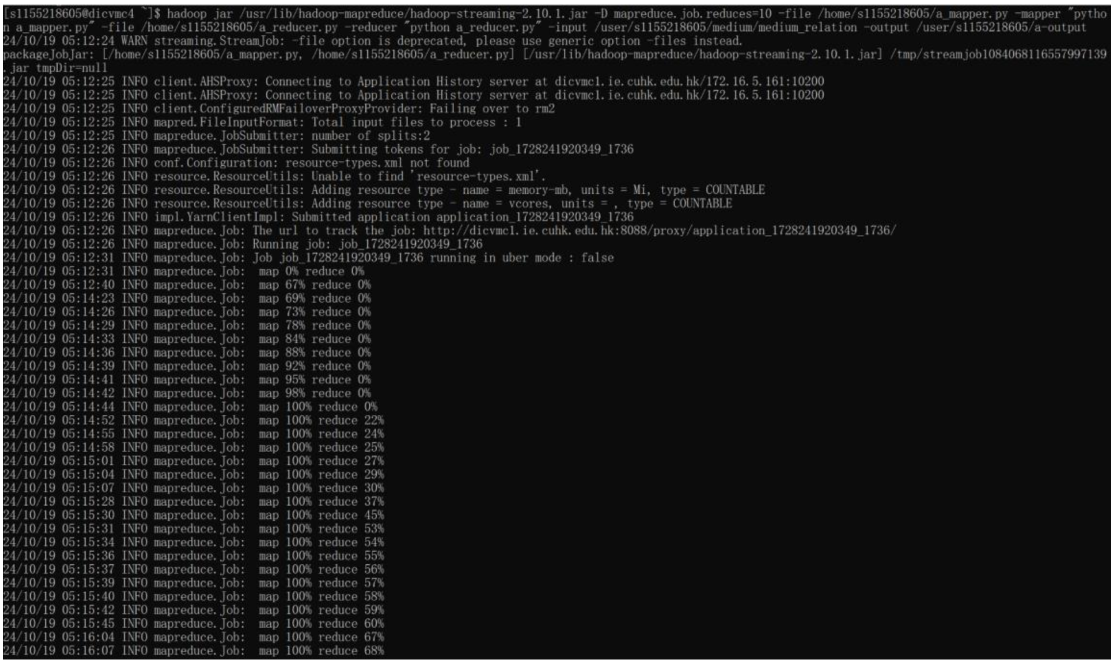
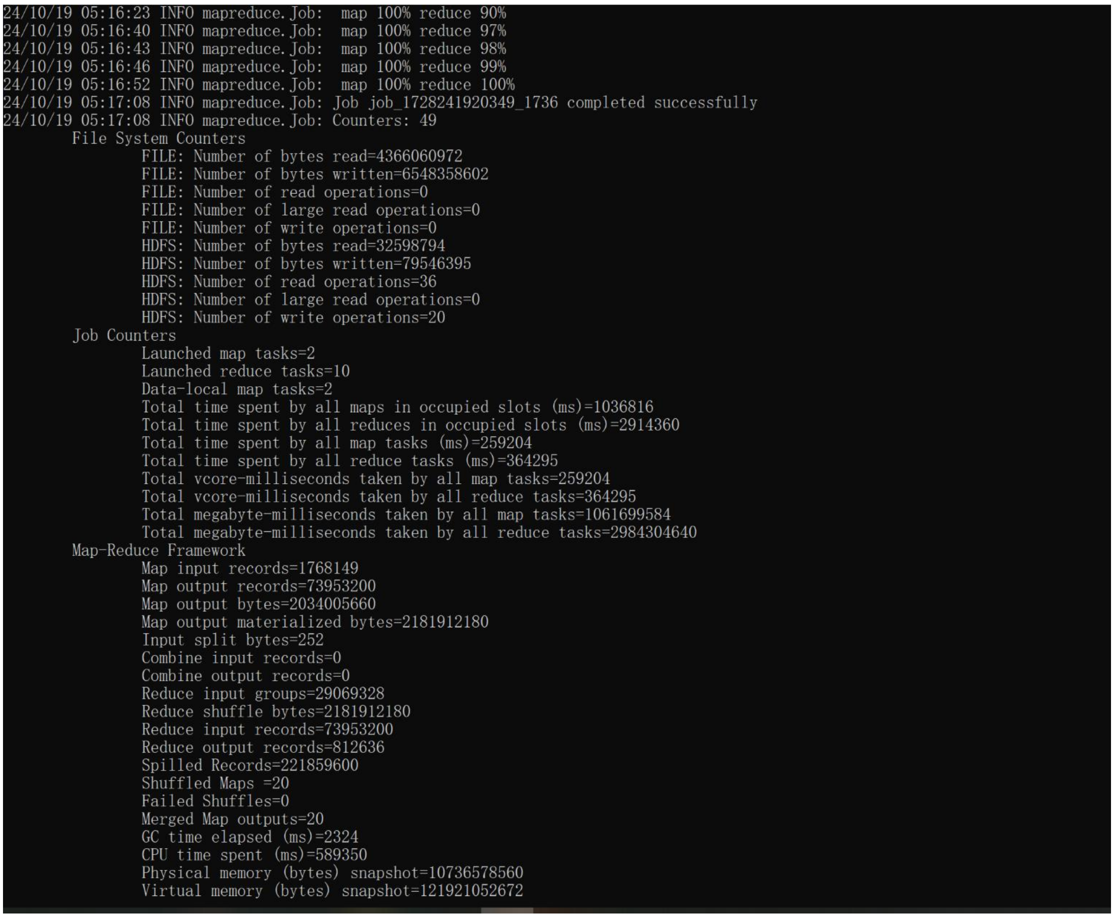
A,

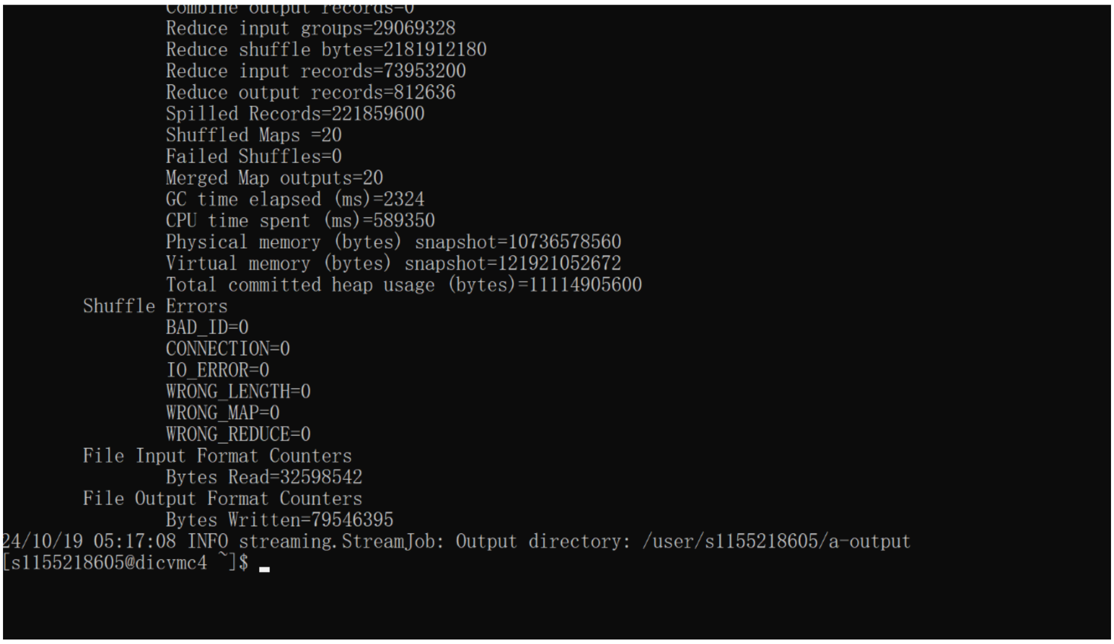
For EVERY website, recommend the website with the maximal number of common referrees in the medium-sized dataset. If multiple websites share the same number, pick the one with the smallest ID.



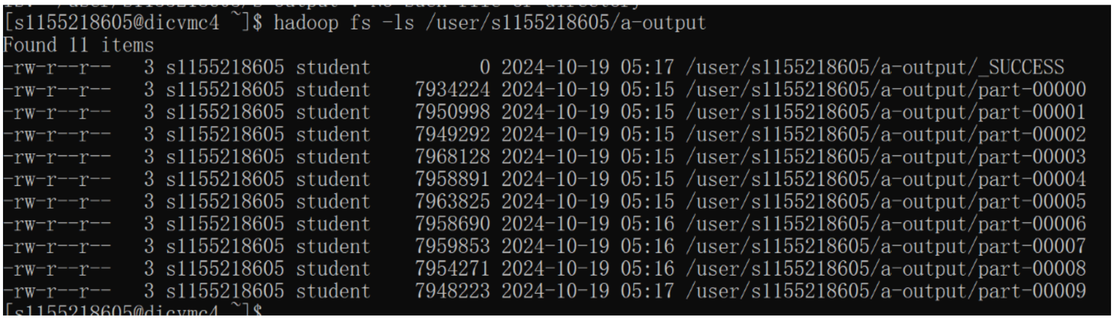
```
[s1155218605@dicvmc4 ~]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming-2.10.1.jar -D mapreduce.job.reduces=10 -file /home/s1155218605/a_mapper.py -mapper "pytho
n a_mapper.py" -file /home/s1155218605/a_reducer.py -reducer "python a_reducer.py" -input /user/s1155218605/medium/medium_relation -output /user/s1155218605/a-output
24/10/19 05:12:24 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
packageJobJar: [/home/s1155218605/a_mapper.py, /home/s1155218605/a_reducer.py] [/usr/lib/hadoop-mapreduce/hadoop-streaming-2.10.1.jar] /tmp/streamjob1084068116557997139
.jar tmpDir=null
24/10/19 05:12:25 INFO client.AHSProxy: Connecting to Application History server at dicvmc1.ie.cuhk.edu.hk/172.16.5.161:10200
24/10/19 05:12:25 INFO client.AHSProxy: Connecting to Application History server at dicvmc1.ie.cuhk.edu.hk/172.16.5.161:10200
24/10/19 05:12:25 INFO client.ConfiguredRMFailoverProxyProvider: Failing over to rm2
24/10/19 05:12:25 INFO mapred.FileInputFormat: Total input files to process : 1
24/10/19 05:12:25 INFO mapreduce.JobSubmitter: number of splits:2
24/10/19 05:12:26 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1728241920349_1736
24/10/19 05:12:26 INFO conf.Configuration: resource-types.xml not found
24/10/19 05:12:26 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
24/10/19 05:12:26 INFO resource.ResourceUtils: Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE
24/10/19 05:12:26 INFO resource.ResourceUtils: Adding resource type - name = vcores, units = , type = COUNTABLE
24/10/19 05:12:26 INFO impl.YarnClientImpl: Submitted application application_1728241920349_1736
24/10/19 05:12:26 INFO mapreduce.Job: The url to track the job: http://dicvmc1.ie.cuhk.edu.hk:8088/proxy/application_1728241920349_1736/
24/10/19 05:12:26 INFO mapreduce.Job: Running job: job_1728241920349_1736
24/10/19 05:12:31 INFO mapreduce.Job: Job job_1728241920349_1736 running in uber mode : false
24/10/19 05:12:31 INFO mapreduce.Job:  map 0% reduce 0%
24/10/19 05:12:40 INFO mapreduce.Job:  map 67% reduce 0%
24/10/19 05:14:23 INFO mapreduce.Job:  map 69% reduce 0%
24/10/19 05:14:26 INFO mapreduce.Job:  map 73% reduce 0%
24/10/19 05:14:29 INFO mapreduce.Job:  map 78% reduce 0%
24/10/19 05:14:33 INFO mapreduce.Job:  map 84% reduce 0%
24/10/19 05:14:36 INFO mapreduce.Job:  map 88% reduce 0%
24/10/19 05:14:39 INFO mapreduce.Job:  map 92% reduce 0%
24/10/19 05:14:41 INFO mapreduce.Job:  map 95% reduce 0%
24/10/19 05:14:42 INFO mapreduce.Job:  map 98% reduce 0%
24/10/19 05:14:44 INFO mapreduce.Job:  map 100% reduce 0%
24/10/19 05:14:52 INFO mapreduce.Job:  map 100% reduce 22%
24/10/19 05:14:55 INFO mapreduce.Job:  map 100% reduce 24%
24/10/19 05:14:58 INFO mapreduce.Job:  map 100% reduce 25%
24/10/19 05:15:01 INFO mapreduce.Job:  map 100% reduce 27%
24/10/19 05:15:04 INFO mapreduce.Job:  map 100% reduce 29%
24/10/19 05:15:07 INFO mapreduce.Job:  map 100% reduce 30%
24/10/19 05:15:28 INFO mapreduce.Job:  map 100% reduce 37%
24/10/19 05:15:30 INFO mapreduce.Job:  map 100% reduce 45%
24/10/19 05:15:31 INFO mapreduce.Job:  map 100% reduce 53%
24/10/19 05:15:34 INFO mapreduce.Job:  map 100% reduce 54%
24/10/19 05:15:36 INFO mapreduce.Job:  map 100% reduce 55%
24/10/19 05:15:37 INFO mapreduce.Job:  map 100% reduce 56%
24/10/19 05:15:39 INFO mapreduce.Job:  map 100% reduce 57%
24/10/19 05:15:40 INFO mapreduce.Job:  map 100% reduce 58%
24/10/19 05:15:42 INFO mapreduce.Job:  map 100% reduce 59%
24/10/19 05:15:45 INFO mapreduce.Job:  map 100% reduce 60%
24/10/19 05:16:04 INFO mapreduce.Job:  map 100% reduce 67%
24/10/19 05:16:07 INFO mapreduce.Job:  map 100% reduce 68%
```



```
24/10/19 05:16:23 INFO mapreduce.Job:  map 100% reduce 90%
24/10/19 05:16:40 INFO mapreduce.Job:  map 100% reduce 97%
24/10/19 05:16:43 INFO mapreduce.Job:  map 100% reduce 98%
24/10/19 05:16:46 INFO mapreduce.Job:  map 100% reduce 99%
24/10/19 05:16:52 INFO mapreduce.Job:  map 100% reduce 100%
24/10/19 05:17:08 INFO mapreduce.Job: Job job_1728241920349_1736 completed successfully
24/10/19 05:17:08 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=4366060972
                FILE: Number of bytes written=6548358602
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=32598794
                HDFS: Number of bytes written=79546395
                HDFS: Number of read operations=36
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=20
        Job Counters
                Launched map tasks=2
                Launched reduce tasks=10
                Data-local map tasks=2
                Total time spent by all maps in occupied slots (ms)=1036816
                Total time spent by all reduces in occupied slots (ms)=2914360
                Total time spent by all map tasks (ms)=259204
                Total time spent by all reduce tasks (ms)=364295
                Total vcore-milliseconds taken by all map tasks=259204
                Total vcore-milliseconds taken by all reduce tasks=364295
                Total megabyte-milliseconds taken by all map tasks=1061699584
                Total megabyte-milliseconds taken by all reduce tasks=2984304640
        Map-Reduce Framework
                Map input records=1768149
                Map output records=73953200
                Map output bytes=2034005660
                Map output materialized bytes=2181912180
                Input split bytes=252
                Combine input records=0
                Combine output records=0
                Reduce input groups=29069328
                Reduce shuffle bytes=2181912180
                Reduce input records=73953200
                Reduce output records=812636
                Spilled Records=221859600
                Shuffled Maps =20
                Failed Shuffles=0
                Merged Map outputs=20
                GC time elapsed (ms)=2324
                CPU time spent (ms)=589350
                Physical memory (bytes) snapshot=10736578560
                Virtual memory (bytes) snapshot=121921052672
```

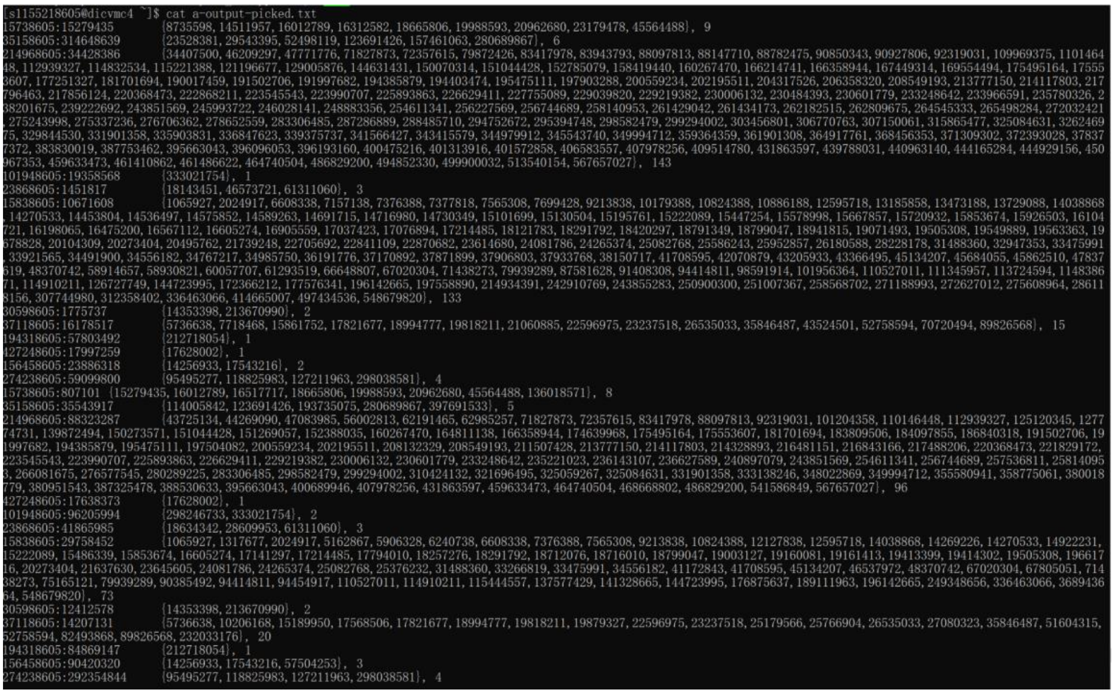


```
                Combine output records=0
                Reduce input groups=29069328
                Reduce shuffle bytes=2181912180
                Reduce input records=73953200
                Reduce output records=812636
                Spilled Records=221859600
                Shuffled Maps =20
                Failed Shuffles=0
                Merged Map outputs=20
                GC time elapsed (ms)=2324
                CPU time spent (ms)=589350
                Physical memory (bytes) snapshot=10736578560
                Virtual memory (bytes) snapshot=121921052672
                Total committed heap usage (bytes)=11114905600
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=32598542
        File Output Format Counters
                Bytes Written=79546395
24/10/19 05:17:08 INFO streaming.StreamJob: Output directory: /user/s1155218605/a-output
[s1155218605@dicvmc4 ~]$ 
```

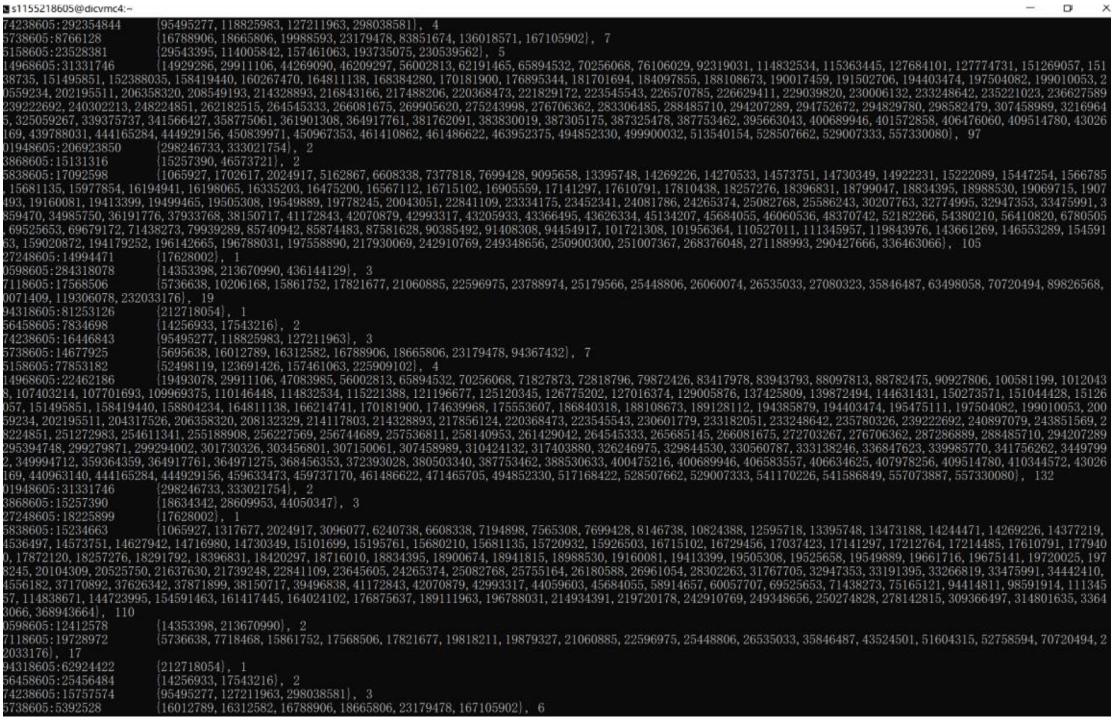


```
[s1155218605@dicvmc4 ~]$ hadoop fs -ls /user/s1155218605/a-output
Found 11 items
-rw-r--r--   3 s1155218605 student          0 2024-10-19 05:17 /user/s1155218605/a-output/_SUCCESS
-rw-r--r--   3 s1155218605 student    7934224 2024-10-19 05:15 /user/s1155218605/a-output/part-00000
-rw-r--r--   3 s1155218605 student    7950998 2024-10-19 05:15 /user/s1155218605/a-output/part-00001
-rw-r--r--   3 s1155218605 student    7949292 2024-10-19 05:15 /user/s1155218605/a-output/part-00002
-rw-r--r--   3 s1155218605 student    7968128 2024-10-19 05:15 /user/s1155218605/a-output/part-00003
-rw-r--r--   3 s1155218605 student    7958891 2024-10-19 05:15 /user/s1155218605/a-output/part-00004
-rw-r--r--   3 s1155218605 student    7963825 2024-10-19 05:15 /user/s1155218605/a-output/part-00005
-rw-r--r--   3 s1155218605 student    7958690 2024-10-19 05:16 /user/s1155218605/a-output/part-00006
-rw-r--r--   3 s1155218605 student    7959853 2024-10-19 05:16 /user/s1155218605/a-output/part-00007
-rw-r--r--   3 s1155218605 student    7954271 2024-10-19 05:16 /user/s1155218605/a-output/part-00008
-rw-r--r--   3 s1155218605 student    7948223 2024-10-19 05:17 /user/s1155218605/a-output/part-00009
[s1155218605@dicvmc4 ~]$
```

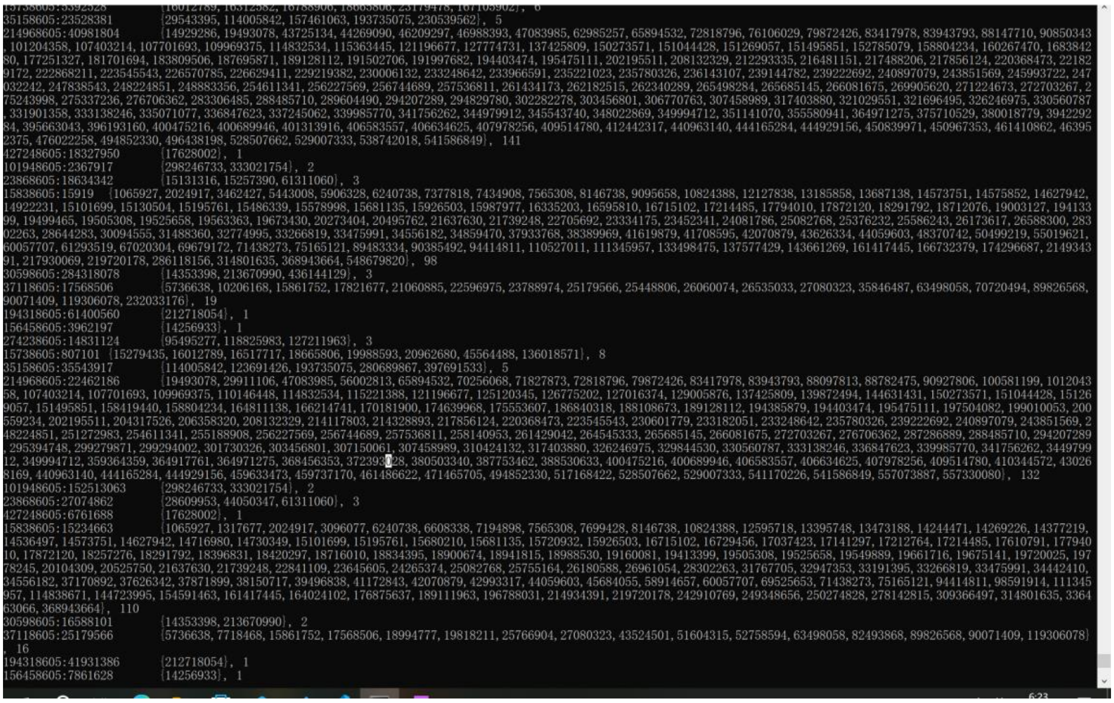
Write python code pick_num.py display numbers



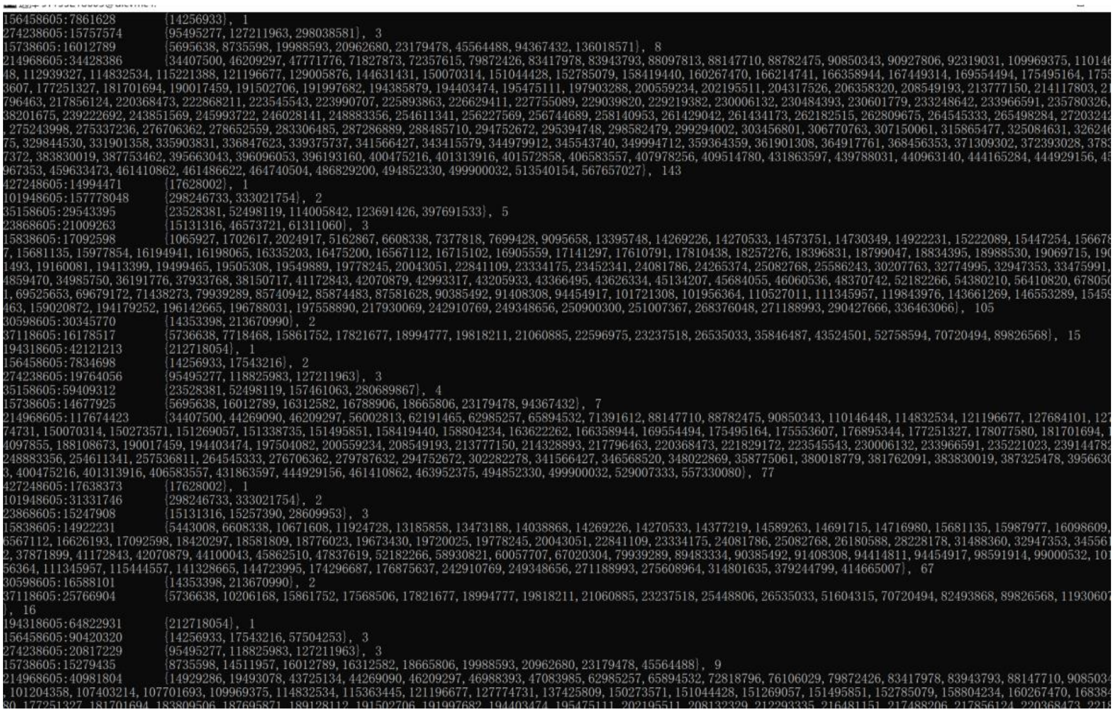
```
[s1155218605@dicvmc4 ~]$ cat a-output-picked.txt
15738605:15279435        {8735598, 14511957, 16012789, 16312582, 18665806, 19988593, 20962680, 23179478, 45564488},  9
35158605:314648639       {23528381, 29543395, 52498119, 123691426, 157461063, 280689867},  6
214968605:34428386       {34407500, 46209297, 47771776, 71827873, 72357615, 79872426, 83417978, 83943793, 88097813, 88147710, 88782475, 90850343, 90927806, 92319031, 109969375, 1101464
48, 112939327, 114832534, 115221388, 121196677, 129005876, 144631431, 150070314, 151044428, 152785079, 158419440, 160267470, 166214741, 166358944, 167449314, 169554494, 175495164, 17555
3607, 177251327, 181701694, 190017459, 191502706, 191997682, 194385879, 194403474, 195475111, 197903288, 200559234, 202195511, 204317526, 206358320, 208549193, 213777150, 214117803, 217
796463, 217856124, 220368473, 222868211, 223545543, 223990707, 225893863, 226629411, 227755089, 229039820, 229219382, 230006132, 230484393, 230601779, 233248642, 233966591, 235780326, 2
38201675, 239222692, 243851569, 245993722, 246028141, 248883356, 254611341, 256227569, 256744689, 258140953, 261429042, 261434173, 262182515, 262809675, 264545333, 265498284, 272032421
, 275243998, 275337236, 276706362, 278652559, 283306485, 287286889, 288485710, 294752672, 295394748, 298582479, 299294002, 303456801, 306770763, 307150061, 315865477, 325084631, 3262469
75, 329844530, 331901358, 335903831, 336847623, 339375737, 341566427, 343415579, 344979912, 345543740, 349994712, 359364359, 361901308, 364917761, 368456353, 371309302, 372393028, 37837
7372, 383830019, 387753462, 395663043, 396096053, 396193160, 400475216, 401313916, 401572858, 406583557, 407978256, 409514780, 431863597, 439788031, 440963140, 444165284, 444929156, 450
967353, 459633473, 461410862, 461486622, 464740504, 486829200, 494852330, 499900032, 513540154, 567657027},  143
101948605:19358568       {333021754},  1
23868605:1451817         {18143451, 46573721, 61311060},  3
15838605:10671608        {1065927, 2024917, 6608338, 7157138, 7376388, 7377818, 7565308, 7699428, 9213838, 10179388, 10824388, 10886188, 12595718, 13185858, 13473188, 13729088, 14038868
, 14270533, 14453804, 14536497, 14575852, 14589263, 14691715, 14716980, 14730349, 15101699, 15130504, 15195761, 15222089, 15447254, 15578998, 15667857, 15720932, 15853674, 15926503, 16104
721, 16198065, 16475200, 16567112, 16605274, 16905559, 17037423, 17076894, 17214485, 18121783, 18291792, 18420297, 18791349, 18799047, 18941815, 19071493, 19505308, 19549889, 19563363, 19
678828, 20104309, 20273404, 20495762, 21739248, 22705692, 22841109, 22870682, 23614680, 24081786, 24265374, 25082768, 25586243, 25952857, 26180588, 28228178, 31488360, 32947353, 33475991
, 33921565, 34491900, 34556182, 34767217, 34985750, 36191776, 37170892, 37871899, 37906803, 37933768, 38150717, 41708595, 42070879, 43205933, 43366495, 45134207, 45684055, 45862510, 47837
619, 48370742, 58914657, 58930821, 60057707, 61293519, 66648807, 67020304, 71438273, 79939289, 87581628, 91408308, 94414811, 98591914, 101956364, 110527011, 111345957, 113724594, 1148386
71, 114910211, 126727749, 144723995, 172366212, 177576341, 196142665, 197558890, 214934391, 242910769, 243855283, 250900300, 251007367, 258568702, 271188993, 272627012, 275608964, 28611
8156, 307744980, 312358402, 336463066, 414665007, 497434536, 548679820},  133
30598605:1775737         {14353398, 213670990},  2
37118605:16178517        {5736638, 7718468, 15861752, 17821677, 18994777, 19818211, 21060885, 22596975, 23237518, 26535033, 35846487, 43524501, 52758594, 70720494, 89826568},  15
194318605:57803492       {212718054},  1
427248605:17997259       {17628002},  1
156458605:23886318       {14256933, 17543216},  2
274238605:59099800       {95495277, 118825983, 127211963, 298038581},  4
15738605:807101  {15279435, 16012789, 16517717, 18665806, 19988593, 20962680, 45564488, 136018571},  8
35158605:35543917        {114005842, 123691426, 193735075, 280689867, 397691533},  5
214968605:88323287       {43725134, 44269090, 47083985, 56002813, 62191465, 62985257, 71827873, 72357615, 83417978, 88097813, 92319031, 101204358, 110146448, 112939327, 125120345, 1277
74731, 139872494, 150273571, 151044428, 151269057, 152388035, 160267470, 164811138, 166358944, 174639968, 175495164, 175553607, 181701694, 183809506, 184097855, 186840318, 191502706, 19
1997682, 194385879, 195475111, 197504082, 200559234, 202195511, 208132329, 208549193, 211507428, 213777150, 214117803, 214328893, 216481151, 216843166, 217488206, 220368473, 221829172,
223545543, 223990707, 225893863, 226629411, 229219382, 230006132, 230601779, 233248642, 235221023, 236143107, 236627589, 240897079, 243851569, 254611341, 256744689, 257536811, 25814095
3, 266081675, 276577545, 280289225, 283306485, 298582479, 299294002, 310424132, 321696495, 325059267, 325084631, 331901358, 333138246, 348022869, 349994712, 355580941, 358775061, 380018
779, 380951543, 387325478, 388530633, 395663043, 400689946, 407978256, 431863597, 459633473, 464740504, 468668802, 486829200, 541586849, 567657027},  96
427248605:17638373       {17628002},  1
101948605:96205994       {298246733, 333021754},  2
23868605:41865985        {18634342, 28609953, 61311060},  3
15838605:29758452        {1065927, 1317677, 2024917, 5162867, 5906328, 6240738, 6608338, 7376388, 7565308, 9213838, 10824388, 12127838, 12595718, 14038868, 14269226, 14270533, 14922231,
15222089, 15486339, 15853674, 16605274, 17141297, 17214485, 17794010, 18257276, 18291792, 18712076, 18716010, 18799047, 19003127, 19160081, 19161413, 19413399, 19414302, 19505308, 196617
16, 20273404, 21637630, 23645605, 24081786, 24265374, 25082768, 25376232, 31488360, 33266819, 33475991, 34556182, 41172843, 41708595, 45134207, 46537972, 48370742, 67020304, 67805051, 714
38273, 75165121, 79939289, 90385492, 94414811, 94454917, 110527011, 114910211, 115444557, 137577429, 141328665, 144723995, 176875637, 189111963, 196142665, 249348656, 336463066, 3689436
64, 548679820},  73
30598605:12412578        {14353398, 213670990},  2
37118605:14207131        {5736638, 10206168, 15189950, 17568506, 17821677, 18994777, 19818211, 19879327, 22596975, 23237518, 25179566, 25766904, 26535033, 27080323, 35846487, 51604315,
52758594, 82493868, 89826568, 232033176},  20
194318605:84869147       {212718054},  1
156458605:90420320       {14256933, 17543216, 57504253},  3
274238605:292354844      {95495277, 118825983, 127211963, 298038581},  4
```



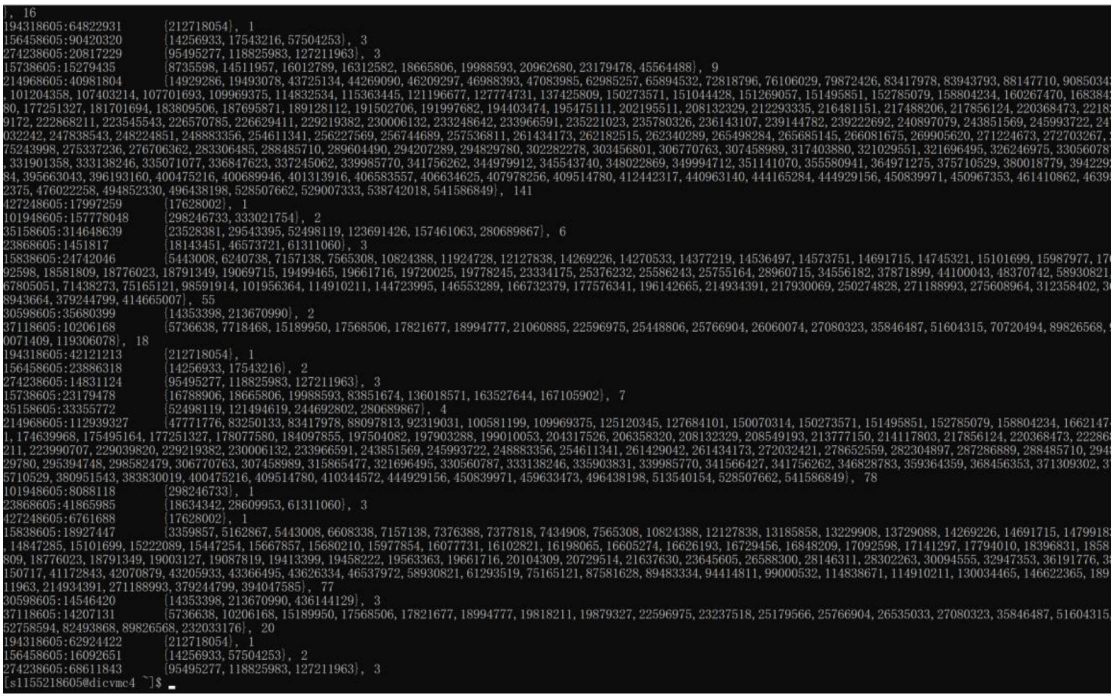
```
74238605:292354844      {95495277, 118825983, 127211963, 298038581}, 4
5738605:8766128         {16788906, 18665806, 19988593, 23179478, 83851674, 136018571, 167105902}, 7
5158605:23528381        {29543395, 114005842, 157461063, 193735075, 230539562}, 5
14968605:31331746       {14929286, 29911106, 44269090, 46209297, 56002813, 62191465, 65894532, 70256068, 76106029, 92319031, 114832534, 115363445, 127684101, 127774731, 151269057, 151
38735, 151495851, 152388035, 158419440, 160267470, 164811138, 168384280, 170181900, 176895344, 181701694, 184097855, 188108673, 190017459, 191502706, 194403474, 197504082, 199010053, 2
0559234, 202195511, 206358320, 208549193, 214328893, 216843166, 217488206, 220368473, 221829172, 223545543, 226570785, 226629411, 229039820, 230006132, 233248642, 235221023, 236627589
239222692, 240302213, 248224851, 262182515, 264545333, 266081675, 269905620, 275243998, 276706362, 283306485, 288485710, 294207289, 294752672, 294829780, 298582479, 307458989, 3216964
5, 325059267, 339375737, 341566427, 358775061, 361901308, 364917761, 381762091, 383830019, 387305175, 387325478, 387753462, 395663043, 400689946, 401572858, 406476060, 409514780, 43026
169, 439788031, 444165284, 444929156, 450839971, 450967353, 461410862, 461486622, 463952375, 494852330, 499900032, 513540154, 528507662, 529007333, 557330080}, 97
01948605:206923850      {298246733, 333021754}, 2
3868605:15131316        {15257390, 46573721}, 2
5838605:17092598        {1065927, 1702617, 2024917, 5162867, 6608338, 7377818, 7699428, 9095658, 13395748, 14269226, 14270533, 14573751, 14730349, 14922231, 15222089, 15447254, 1566785
, 15681135, 15977854, 16194941, 16198065, 16335203, 16475200, 16567112, 16715102, 16905559, 17141297, 17610791, 17810438, 18257276, 18396831, 18799047, 18834395, 18988530, 19069715, 1907
493, 19160081, 19413399, 19499465, 19505308, 19549889, 19778245, 20043051, 22841109, 23334175, 23452341, 24081786, 24265374, 25082768, 25586243, 30207763, 32774995, 32947353, 33475991, 3
859470, 34985750, 36191776, 37933768, 38150717, 41172843, 42070879, 42993317, 43205933, 43366495, 43626334, 45134207, 45684055, 46060536, 48370742, 52182266, 54380210, 56410820, 6780505
, 69525653, 69679172, 71438273, 79939289, 85740942, 85874483, 87581628, 90385492, 91408308, 94454917, 101721308, 101956364, 110527011, 111345957, 119843976, 143661269, 146553289, 154591
63, 159020872, 194179252, 196142665, 196788031, 197558890, 217930069, 242910769, 249348656, 250900300, 251007367, 268376048, 271188993, 290427666, 336463066}, 105
27248605:14994471       {17628002}, 1
0598605:284318078       {14353398, 213670990, 436144129}, 3
7118605:17568506        {5736638, 10206168, 15861752, 17821677, 21060885, 22596975, 23788974, 25179566, 25448806, 26060074, 26535033, 27080323, 35846487, 63498058, 70720494, 89826568,
0071409, 119306078, 232033176}, 19
94318605:81253126       {212718054}, 1
56458605:7834698        {14256933, 17543216}, 2
74238605:16446843       {95495277, 118825983, 127211963}, 3
5738605:14677925        {5695638, 16012789, 16312582, 16788906, 18665806, 23179478, 94367432}, 7
5158605:77853182        {52498119, 123691426, 157461063, 225909102}, 4
14968605:22462186       {19493078, 29911106, 47083985, 56002813, 65894532, 70256068, 71827873, 72818796, 79872426, 83417978, 83943793, 88097813, 88782475, 90927806, 100581199, 1012043
8, 107403214, 107701693, 109969375, 110146448, 114832534, 115221388, 121196677, 125120345, 126775202, 127016374, 129005876, 137425809, 139872494, 144631431, 150273571, 151044428, 15126
057, 151495851, 158419440, 158804234, 164811138, 166214741, 170181900, 174639968, 175553607, 186840318, 188108673, 189128112, 194385879, 194403474, 195475111, 197504082, 199010053, 200
59234, 202195511, 204317526, 206358320, 208132329, 214117803, 214328893, 217856124, 220368473, 223545543, 230601779, 233182051, 233248642, 235780326, 239222692, 240897079, 243851569, 2
8224851, 251272983, 254611341, 255188908, 256227569, 256744689, 257536811, 258140953, 261429042, 264545333, 265685145, 266081675, 272703267, 276706362, 287286889, 288485710, 294207289
295394748, 299279871, 299294002, 301730326, 303456801, 307150061, 307458989, 310424132, 317403880, 326246975, 329844530, 330560787, 333138246, 336847623, 339985770, 341756262, 3449799
2, 349994712, 359364359, 364917761, 364971275, 368456353, 372393028, 380503340, 387753462, 388530633, 400475216, 400689946, 406583557, 406634625, 407978256, 409514780, 410344572, 43026
169, 440963140, 444165284, 444929156, 459633473, 459737170, 461486622, 471465705, 494852330, 517168422, 528507662, 529007333, 541170226, 541586849, 557073887, 557330080}, 132
01948605:31331746       {298246733, 333021754}, 2
3868605:15257390        {18634342, 28609953, 44050347}, 3
27248605:18225899       {17628002}, 1
5838605:15234663        {1065927, 1317677, 2024917, 3096077, 6240738, 6608338, 7194898, 7565308, 7699428, 8146738, 10824388, 12595718, 13395748, 13473188, 14244471, 14269226, 14377219,
4536497, 14573751, 14627942, 14716980, 14730349, 15101699, 15195761, 15680210, 15681135, 15720932, 15926503, 16715102, 16729456, 17037423, 17141297, 17212764, 17214485, 17610791, 177940
0, 17872120, 18257276, 18291792, 18396831, 18420297, 18716010, 18834395, 18900674, 18941815, 18988530, 19160081, 19413399, 19505308, 19525658, 19549889, 19661716, 19675141, 19720025, 197
8245, 20104309, 20525750, 21637630, 21739248, 22841109, 23645605, 24265374, 25082768, 25755164, 26180588, 26961054, 28302263, 31767705, 32947353, 33191395, 33266819, 33475991, 34442410,
4556182, 37170892, 37626342, 37871899, 38150717, 39496838, 41172843, 42070879, 42993317, 44059603, 45684055, 58914657, 60057707, 69525653, 71438273, 75165121, 94414811, 98591914, 111345
57, 114838671, 144723995, 154591463, 161417445, 164024102, 176875637, 189111963, 196788031, 214934391, 219720178, 242910769, 249348656, 250274828, 278142815, 309366497, 314801635, 3364
3066, 368943664}, 110
0598605:12412578        {14353398, 213670990}, 2
7118605:19728972        {5736638, 7718468, 15861752, 17568506, 17821677, 19818211, 19879327, 21060885, 22596975, 25448806, 26535033, 35846487, 43524501, 51604315, 52758594, 70720494, 2
2033176}, 17
94318605:62924422       {212718054}, 1
56458605:25456484       {14256933, 17543216}, 2
74238605:15757574       {95495277, 127211963, 298038581}, 3
5738605:5392528         {16012789, 16312582, 16788906, 18665806, 23179478, 167105902}, 6
```



```
15738605:5392528        {16012789, 16312582, 16788906, 18665806, 23179478, 167105902}, 6
35158605:23528381       {29543395, 114005842, 157461063, 193735075, 230539562}, 5
214968605:40981804      {14929286, 19493078, 43725134, 44269090, 46209297, 46988393, 47083985, 62985257, 65894532, 72818796, 76106029, 79872426, 83417978, 83943793, 88147710, 90850343
, 101204358, 107403214, 107701693, 109969375, 114832534, 115363445, 121196677, 127774731, 137425809, 150273571, 151044428, 151269057, 151495851, 152785079, 158804234, 160267470, 1683842
80, 177251327, 181701694, 183809506, 187695871, 189128112, 191502706, 191997682, 194403474, 195475111, 202195511, 208132329, 212293335, 216481151, 217488206, 217856124, 220368473, 22182
9172, 222868211, 223545543, 226570785, 226629411, 229219382, 230006132, 233248642, 233966591, 235221023, 235780326, 236143107, 239144782, 239222692, 240897079, 243851569, 245993722, 247
032242, 247838543, 248224851, 248883356, 254611341, 256227569, 256744689, 257536811, 261434173, 262182515, 262340289, 265498284, 265685145, 266081675, 269905620, 271224673, 272703267, 2
75243998, 275337236, 276706362, 283306485, 288485710, 289604490, 294207289, 294829780, 302282278, 303456801, 306770763, 307458989, 317403880, 321029551, 321696495, 326246975, 330560787
, 331901358, 333138246, 335071077, 336847623, 337245062, 339985770, 341756262, 344979912, 345543740, 348022869, 349994712, 351141070, 355580941, 364971275, 375710529, 380018779, 3942292
84, 395663043, 396193160, 400475216, 400689946, 401313916, 406583557, 406634625, 407978256, 409514780, 412442317, 440963140, 444165284, 444929156, 450839971, 450967353, 461410862, 46395
2375, 476022258, 494852330, 496438198, 528507662, 529007333, 538742018, 541586849}, 141
427248605:18327950      {17628002}, 1
101948605:2367917       {298246733, 333021754}, 2
23868605:18634342       {15131316, 15257390, 61311060}, 3
15838605:15919  {1065927, 2024917, 3462427, 5443008, 5906328, 6240738, 7377818, 7434908, 7565308, 8146738, 9095658, 10824388, 12127838, 13185858, 13687138, 14573751, 14575852, 14627942,
14922231, 15101699, 15130504, 15195761, 15486339, 15578998, 15681135, 15926503, 15987977, 16335203, 16595810, 16715102, 17214485, 17794010, 17872120, 18291792, 18712076, 19003127, 194133
99, 19499465, 19505308, 19525658, 19563363, 19673430, 20273404, 20495762, 21637630, 21739248, 22705692, 23334175, 23452341, 24081786, 25082768, 25376232, 25586243, 26173617, 26588300, 283
02263, 28644283, 30094555, 31488360, 32774995, 33266819, 33475991, 34556182, 34859470, 37933768, 38389969, 41619879, 41708595, 42070879, 43626334, 44059603, 48370742, 50499219, 55019621,
60057707, 61293519, 67020304, 69679172, 71438273, 75165121, 89483334, 90385492, 94414811, 110527011, 111345957, 133498475, 137577429, 143661269, 161417445, 166732379, 174296687, 2149343
91, 217930069, 219720178, 286118156, 314801635, 368943664, 548679820}, 98
30598605:284318078      {14353398, 213670990, 436144129}, 3
37118605:17568506       {5736638, 10206168, 15861752, 17821677, 21060885, 22596975, 23788974, 25179566, 25448806, 26060074, 26535033, 27080323, 35846487, 63498058, 70720494, 89826568,
90071409, 119306078, 232033176}, 19
194318605:61400560      {212718054}, 1
156458605:3962197       {14256933}, 1
274238605:14831124      {95495277, 118825983, 127211963}, 3
15738605:807101 {15279435, 16012789, 16517717, 18665806, 19988593, 20962680, 45564488, 136018571}, 8
35158605:35543917       {114005842, 123691426, 193735075, 280689867, 397691533}, 5
214968605:22462186      {19493078, 29911106, 47083985, 56002813, 65894532, 70256068, 71827873, 72818796, 79872426, 83417978, 83943793, 88097813, 88782475, 90927806, 100581199, 1012043
58, 107403214, 107701693, 109969375, 110146448, 114832534, 115221388, 121196677, 125120345, 126775202, 127016374, 129005876, 137425809, 139872494, 144631431, 150273571, 151044428, 15126
9057, 151495851, 158419440, 158804234, 164811138, 166214741, 170181900, 174639968, 175553607, 186840318, 188108673, 189128112, 194385879, 194403474, 195475111, 197504082, 199010053, 200
559234, 202195511, 204317526, 206358320, 208132329, 214117803, 214328893, 217856124, 220368473, 223545543, 230601779, 233182051, 233248642, 235780326, 239222692, 240897079, 243851569, 2
48224851, 251272983, 254611341, 255188908, 256227569, 256744689, 257536811, 258140953, 261429042, 264545333, 265685145, 266081675, 272703267, 276706362, 287286889, 288485710, 294207289
, 295394748, 299279871, 299294002, 301730326, 303456801, 307150061, 307458989, 310424132, 317403880, 326246975, 329844530, 330560787, 333138246, 336847623, 339985770, 341756262, 3449799
12, 349994712, 359364359, 364917761, 364971275, 368456353, 372393028, 380503340, 387753462, 388530633, 400475216, 400689946, 406583557, 406634625, 407978256, 409514780, 410344572, 43026
8169, 440963140, 444165284, 444929156, 459633473, 459737170, 461486622, 471465705, 494852330, 517168422, 528507662, 529007333, 541170226, 541586849, 557073887, 557330080}, 132
101948605:152513063     {298246733, 333021754}, 2
23868605:27074862       {28609953, 44050347, 61311060}, 3
427248605:6761688       {17628002}, 1
15838605:15234663       {1065927, 1317677, 2024917, 3096077, 6240738, 6608338, 7194898, 7565308, 7699428, 8146738, 10824388, 12595718, 13395748, 13473188, 14244471, 14269226, 14377219,
14536497, 14573751, 14627942, 14716980, 14730349, 15101699, 15195761, 15680210, 15681135, 15720932, 15926503, 16715102, 16729456, 17037423, 17141297, 17212764, 17214485, 17610791, 177940
10, 17872120, 18257276, 18291792, 18396831, 18420297, 18716010, 18834395, 18900674, 18941815, 18988530, 19160081, 19413399, 19505308, 19525658, 19549889, 19661716, 19675141, 19720025, 197
78245, 20104309, 20525750, 21637630, 21739248, 22841109, 23645605, 24265374, 25082768, 25755164, 26180588, 26961054, 28302263, 31767705, 32947353, 33191395, 33266819, 33475991, 34442410,
34556182, 37170892, 37626342, 37871899, 38150717, 39496838, 41172843, 42070879, 42993317, 44059603, 45684055, 58914657, 60057707, 69525653, 71438273, 75165121, 94414811, 98591914, 111345
957, 114838671, 144723995, 154591463, 161417445, 164024102, 176875637, 189111963, 196788031, 214934391, 219720178, 242910769, 249348656, 250274828, 278142815, 309366497, 314801635, 3364
63066, 368943664}, 110
30598605:16588101       {14353398, 213670990}, 2
37118605:25179566       {5736638, 7718468, 15861752, 17568506, 18994777, 19818211, 25766904, 27080323, 43524501, 51604315, 52758594, 63498058, 82493868, 89826568, 90071409, 119306078}
, 16
194318605:41931386      {212718054}, 1
156458605:7861628       {14256933}, 1
```



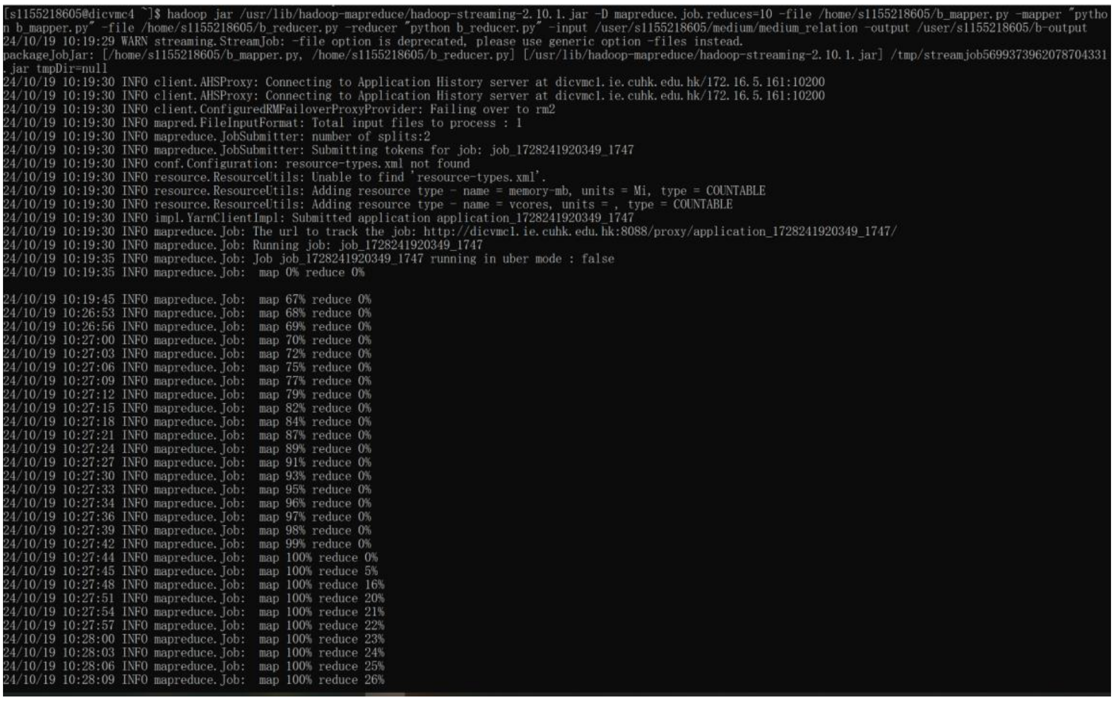
```
156458605:7861628       {14256933}, 1
274238605:15757574      {95495277, 127211963, 298038581}, 3
15738605:16012789       {5695638, 8735598, 19988593, 20962680, 23179478, 45564488, 94367432, 136018571}, 8
214968605:34428386      {34407500, 46209297, 47771776, 71827873, 72357615, 79872426, 83417978, 83943793, 88097813, 88147710, 88782475, 90850343, 90927806, 92319031, 109969375, 110146
48, 112939327, 114832534, 115221388, 121196677, 129005876, 144631431, 150070314, 151044428, 152785079, 158419440, 160267470, 166214741, 166358944, 167449314, 169554494, 175495164, 1755
3607, 177251327, 181701694, 190017459, 191502706, 191997682, 194385879, 194403474, 195475111, 197903288, 200559234, 202195511, 204317526, 206358320, 208549193, 213777150, 214117803, 21
796463, 217856124, 220368473, 222868211, 223545543, 223990707, 225893863, 226629411, 227755089, 229039820, 229219382, 230006132, 230484393, 230601779, 233248642, 233966591, 235780326,
38201675, 239222692, 243851569, 245993722, 246028141, 248883356, 254611341, 256227569, 256744689, 258140953, 261429042, 261434173, 262182515, 262809675, 264545333, 265498284, 27203242
, 275243998, 275337236, 276706362, 278652559, 283306485, 287286889, 288485710, 294752672, 295394748, 298582479, 299294002, 303456801, 306770763, 307150061, 315865477, 325084631, 326246
75, 329844530, 331901358, 335903831, 336847623, 339375737, 341566427, 343415579, 344979912, 345543740, 349994712, 359364359, 361901308, 364917761, 368456353, 371309302, 372393028, 3783
7372, 383830019, 387753462, 395663043, 396096053, 396193160, 400475216, 401313916, 401572858, 406583557, 407978256, 409514780, 431863597, 439788031, 440963140, 444165284, 444929156, 45
967353, 459633473, 461410862, 461486622, 464740504, 486829200, 494852330, 499900032, 513540154, 567657027}, 143
427248605:14994471      {17628002}, 1
101948605:157778048     {298246733, 333021754}, 2
35158605:29543395       {23528381, 52498119, 114005842, 123691426, 397691533}, 5
23868605:21009263       {15131316, 46573721, 61311060}, 3
15838605:17092598       {1065927, 1702617, 2024917, 5162867, 6608338, 7377818, 7699428, 9095658, 13395748, 14269226, 14270533, 14573751, 14730349, 14922231, 15222089, 15447254, 156678
7, 15681135, 15977854, 16194941, 16198065, 16335203, 16475200, 16567112, 16715102, 16905559, 17141297, 17610791, 17810438, 18257276, 18396831, 18799047, 18834395, 18988530, 19069715, 190
1493, 19160081, 19413399, 19499465, 19505308, 19549889, 19778245, 20043051, 22841109, 23334175, 23452341, 24081786, 24265374, 25082768, 25586243, 30207763, 32774995, 32947353, 33475991,
4859470, 34985750, 36191776, 37933768, 38150717, 41172843, 42070879, 42993317, 43205933, 43366495, 43626334, 45134207, 45684055, 46060536, 48370742, 52182266, 54380210, 56410820, 678050
1, 69525653, 69679172, 71438273, 79939289, 85740942, 85874483, 87581628, 90385492, 91408308, 94454917, 101721308, 101956364, 110527011, 111345957, 119843976, 143661269, 146553289, 15459
463, 159020872, 194179252, 196142665, 196788031, 197558890, 217930069, 242910769, 249348656, 250900300, 251007367, 268376048, 271188993, 290427666, 336463066}, 105
30598605:30345770       {14353398, 213670990}, 2
37118605:16178517       {5736638, 7718468, 15861752, 17821677, 18994777, 19818211, 21060885, 22596975, 23237518, 26535033, 35846487, 43524501, 52758594, 70720494, 89826568}, 15
194318605:42121213      {212718054}, 1
156458605:7834698       {14256933, 17543216}, 2
274238605:19764056      {95495277, 118825983, 127211963}, 3
35158605:59409312       {23528381, 52498119, 157461063, 280689867}, 4
15738605:14677925       {5695638, 16012789, 16312582, 16788906, 18665806, 23179478, 94367432}, 7
214968605:117674423     {34407500, 44269090, 46209297, 56002813, 62191465, 62985257, 65894532, 71391612, 88147710, 88782475, 90850343, 110146448, 114832534, 121196677, 127684101, 127
74731, 150070314, 150273571, 151269057, 151338735, 151495851, 158419440, 158804234, 163622262, 166358944, 169554494, 175495164, 175553607, 176895344, 177251327, 178077580, 181701694, 1
4097855, 188108673, 190017459, 194403474, 197504082, 200559234, 208549193, 213777150, 214328893, 217796463, 220368473, 221829172, 223545543, 230006132, 233966591, 235221023, 239144782
248883356, 254611341, 257536811, 264545333, 276706362, 279787632, 294752672, 302282278, 341566427, 346568520, 348022869, 358775061, 380018779, 381762091, 383830019, 387325478, 3956630
3, 400475216, 401313916, 406583557, 431863597, 444929156, 461410862, 463952375, 494852330, 499900032, 529007333, 557330080}, 77
427248605:17638373      {17628002}, 1
101948605:31331746      {298246733, 333021754}, 2
23868605:15247908       {15131316, 15257390, 28609953}, 3
15838605:14922231       {5443008, 6608338, 10671608, 11924728, 13185858, 13473188, 14038868, 14269226, 14270533, 14377219, 14589263, 14691715, 14716980, 15681135, 15987977, 16098609,
6567112, 16626193, 17092598, 18420297, 18581809, 18776023, 19673430, 19720025, 19778245, 20043051, 22841109, 23334175, 24081786, 25082768, 26180588, 28228178, 31488360, 32947353, 345561
2, 37871899, 41172843, 42070879, 44100043, 45862510, 47837619, 52182266, 58930821, 60057707, 67020304, 79939289, 89483334, 90385492, 91408308, 94414811, 94454917, 98591914, 99000532, 101
56364, 111345957, 115444557, 141328665, 144723995, 174296687, 176875637, 242910769, 249348656, 271188993, 275608964, 314801635, 379244799, 414665007}, 67
30598605:16588101       {14353398, 213670990}, 2
37118605:25766904       {5736638, 10206168, 15861752, 17568506, 17821677, 18994777, 19818211, 21060885, 23237518, 25448806, 26535033, 51604315, 70720494, 82493868, 89826568, 1193060
}, 16
194318605:64822931      {212718054}, 1
156458605:90420320      {14256933, 17543216, 57504253}, 3
274238605:20817229      {95495277, 118825983, 127211963}, 3
15738605:15279435       {8735598, 14511957, 16012789, 16312582, 18665806, 19988593, 20962680, 23179478, 45564488}, 9
214968605:40981804      {14929286, 19493078, 43725134, 44269090, 46209297, 46988393, 47083985, 62985257, 65894532, 72818796, 76106029, 79872426, 83417978, 83943793, 88147710, 9085034
, 101204358, 107403214, 107701693, 109969375, 114832534, 115363445, 121196677, 127774731, 137425809, 150273571, 151044428, 151269057, 151495851, 152785079, 158804234, 160267470, 168384
80, 177251327, 181701694, 183809506, 187695871, 189128112, 191502706, 191997682, 194403474, 195475111, 202195511, 208132329, 212293335, 216481151, 217488206, 217856124, 220368473, 2218
```



```
}, 16
194318605:64822931      {212718054}, 1
156458605:90420320      {14256933, 17543216, 57504253}, 3
274238605:20817229      {95495277, 118825983, 127211963}, 3
15738605:15279435       {8735598, 14511957, 16012789, 16312582, 18665806, 19988593, 20962680, 23179478, 45564488}, 9
214968605:40981804      {14929286, 19493078, 43725134, 44269090, 46209297, 46988393, 47083985, 62985257, 65894532, 72818796, 76106029, 79872426, 83417978, 83943793, 88147710, 9085034
, 101204358, 107403214, 107701693, 109969375, 114832534, 115363445, 121196677, 127774731, 137425809, 150273571, 151044428, 151269057, 151495851, 152785079, 158804234, 160267470, 168384
80, 177251327, 181701694, 183809506, 187695871, 189128112, 191502706, 191997682, 194403474, 195475111, 202195511, 208132329, 212293335, 216481151, 217488206, 217856124, 220368473, 2218
9172, 222868211, 223545543, 226570785, 226629411, 229219382, 230006132, 233248642, 233966591, 235221023, 235780326, 236143107, 239144782, 239222692, 240897079, 243851569, 245993722, 24
032242, 247838543, 248224851, 248883356, 254611341, 256227569, 256744689, 257536811, 261434173, 262182515, 262340289, 265498284, 265685145, 266081675, 269905620, 271224673, 272703267,
75243998, 275337236, 276706362, 283306485, 288485710, 289604490, 294207289, 294829780, 302282278, 303456801, 306770763, 307458989, 317403880, 321029551, 321696495, 326246975, 3305607
, 331901358, 333138246, 335071077, 336847623, 337245062, 339985770, 341756262, 344979912, 345543740, 348022869, 349994712, 351141070, 355580941, 364971275, 375710529, 380018779, 394229
84, 395663043, 396193160, 400475216, 400689946, 401313916, 406583557, 406634625, 407978256, 409514780, 412442317, 440963140, 444165284, 444929156, 450839971, 450967353, 461410862, 4639
2375, 476022258, 494852330, 496438198, 528507662, 529007333, 538742018, 541586849}, 141
427248605:17997259      {17628002}, 1
101948605:157778048     {298246733, 333021754}, 2
35158605:314648639      {23528381, 29543395, 52498119, 123691426, 157461063, 280689867}, 6
23868605:1451817        {18143451, 46573721, 61311060}, 3
15838605:24742046       {5443008, 6240738, 7157138, 7565308, 10824388, 11924728, 12127838, 14269226, 14270533, 14377219, 14536497, 14573751, 14691715, 14745321, 15101699, 15987977, 17
92598, 18581809, 18776023, 18791349, 19069715, 19499465, 19661716, 19720025, 19778245, 23334175, 25376232, 25586243, 25755164, 28960715, 34556182, 37871899, 44100043, 48370742, 58930821,
67805051, 71438273, 75165121, 98591914, 101956364, 114910211, 144723995, 146553289, 166732379, 177576341, 196142665, 214934391, 217930069, 250274828, 271188993, 275608964, 312358402, 3
8943664, 379244799, 414665007}, 55
30598605:35680399       {14353398, 213670990}, 2
37118605:10206168       {5736638, 7718468, 15189950, 17568506, 17821677, 18994777, 21060885, 22596975, 25448806, 25766904, 26060074, 27080323, 35846487, 51604315, 70720494, 89826568, 9
0071409, 119306078}, 18
194318605:42121213      {212718054}, 1
156458605:23886318      {14256933, 17543216}, 2
274238605:14831124      {95495277, 118825983, 127211963}, 3
15738605:23179478       {16788906, 18665806, 19988593, 83851674, 136018571, 163527644, 167105902}, 7
35158605:33355772       {52498119, 121494619, 244692802, 280689867}, 4
214968605:112939327     {47771776, 83250133, 83417978, 88097813, 92319031, 100581199, 109969375, 125120345, 127684101, 150070314, 150273571, 151495851, 152785079, 158804234, 1662147
1, 174639968, 175495164, 177251327, 178077580, 184097855, 197504082, 197903288, 199010053, 204317526, 206358320, 208132329, 208549193, 213777150, 214117803, 217856124, 220368473, 22286
211, 223990707, 229039820, 229219382, 230006132, 233966591, 243851569, 245993722, 248883356, 254611341, 261429042, 261434173, 272032421, 278652559, 282304897, 287286889, 288485710, 294
29780, 295394748, 298582479, 306770763, 307458989, 315865477, 321696495, 330560787, 333138246, 335903831, 339985770, 341566427, 341756262, 346828783, 359364359, 368456353, 371309302, 3
5710529, 380951543, 383830019, 400475216, 409514780, 410344572, 444929156, 450839971, 459633473, 496438198, 513540154, 528507662, 541586849}, 78
101948605:8088118       {298246733}, 1
23868605:41865985       {18634342, 28609953, 61311060}, 3
427248605:6761688       {17628002}, 1
15838605:18927447       {3359857, 5162867, 5443008, 6608338, 7157138, 7376388, 7377818, 7434908, 7565308, 10824388, 12127838, 13185858, 13229908, 13729088, 14269226, 14691715, 1479918
, 14847285, 15101699, 15222089, 15447254, 15667857, 15680210, 15977854, 16077731, 16102821, 16198065, 16605274, 16626193, 16729456, 16848209, 17092598, 17141297, 17794010, 18396831, 1858
809, 18776023, 18791349, 19003127, 19087819, 19413399, 19458222, 19563363, 19661716, 20104309, 20729514, 21637630, 23645605, 26588300, 28146311, 28302263, 30094555, 32947353, 36191776, 3
150717, 41172843, 42070879, 43205933, 43366495, 43626334, 46537972, 58930821, 61293519, 75165121, 87581628, 89483334, 94414811, 99000532, 114838671, 114910211, 130034465, 146622365, 189
11963, 214934391, 271188993, 379244799, 394047585}, 77
30598605:14546420       {14353398, 213670990, 436144129}, 3
37118605:14207131       {5736638, 10206168, 15189950, 17568506, 17821677, 18994777, 19818211, 19879327, 22596975, 23237518, 25179566, 25766904, 26535033, 27080323, 35846487, 51604315,
52758594, 82493868, 89826568, 232033176}, 20
194318605:62924422      {212718054}, 1
156458605:16092651      {14256933, 57504253}, 2
274238605:68611843      {95495277, 118825983, 127211963}, 3
[s1155218605@dicvmc4 ~]$ 
```

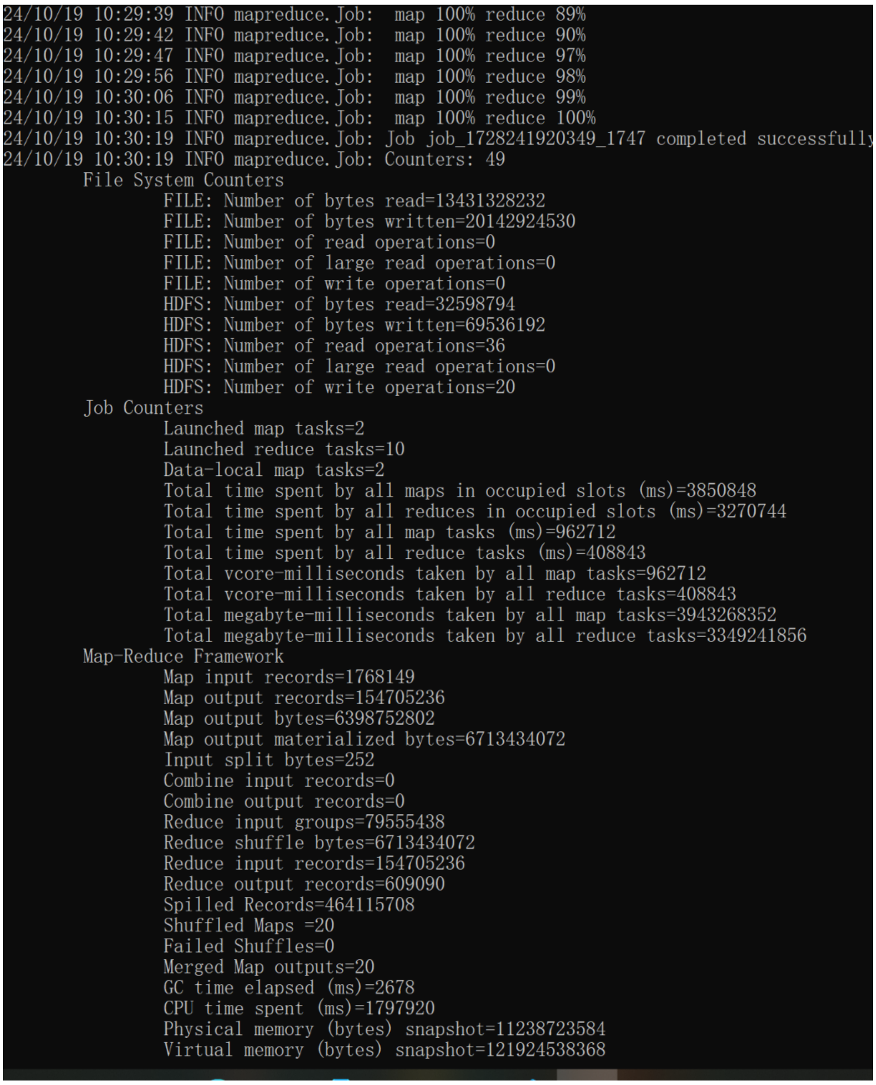
B,

Find the TOP K (K=3) most similar websites of EVERY website as well as their common referrees for the medium-sized dataset [2]. If multiple websites have the same similarity with a particular website, they should all be included in your results. (Still, the total number of records for each website should not exceed K, pick the ones with smaller IDs when there's a tie).

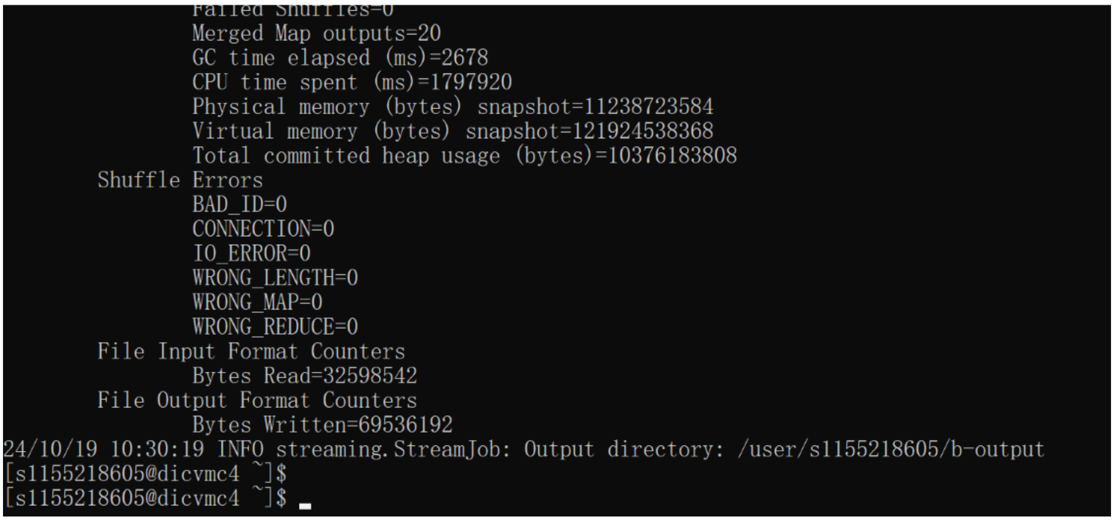


```
[s1155218605@dicvmc4 ~]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming-2.10.1.jar -D mapreduce.job.reduces=10 -file /home/s1155218605/b_mapper.py -mapper "pytho
n b_mapper.py" -file /home/s1155218605/b_reducer.py -reducer "python b_reducer.py" -input /user/s1155218605/medium/medium_relation -output /user/s1155218605/b-output
24/10/19 10:19:29 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
packageJobJar: [/home/s1155218605/b_mapper.py, /home/s1155218605/b_reducer.py] [/usr/lib/hadoop-mapreduce/hadoop-streaming-2.10.1.jar] /tmp/streamjob5699373962078704331
.jar tmpDir=null
24/10/19 10:19:30 INFO client.AHSProxy: Connecting to Application History server at dicvmc1.ie.cuhk.edu.hk/172.16.5.161:10200
24/10/19 10:19:30 INFO client.AHSProxy: Connecting to Application History server at dicvmc1.ie.cuhk.edu.hk/172.16.5.161:10200
24/10/19 10:19:30 INFO client.ConfiguredRMFailoverProxyProvider: Failing over to rm2
24/10/19 10:19:30 INFO mapred.FileInputFormat: Total input files to process : 1
24/10/19 10:19:30 INFO mapreduce.JobSubmitter: number of splits:2
24/10/19 10:19:30 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1728241920349_1747
24/10/19 10:19:30 INFO conf.Configuration: resource-types.xml not found
24/10/19 10:19:30 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
24/10/19 10:19:30 INFO resource.ResourceUtils: Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE
24/10/19 10:19:30 INFO resource.ResourceUtils: Adding resource type - name = vcores, units = , type = COUNTABLE
24/10/19 10:19:30 INFO impl.YarnClientImpl: Submitted application application_1728241920349_1747
24/10/19 10:19:30 INFO mapreduce.Job: The url to track the job: http://dicvmc1.ie.cuhk.edu.hk:8088/proxy/application_1728241920349_1747/
24/10/19 10:19:30 INFO mapreduce.Job: Running job: job_1728241920349_1747
24/10/19 10:19:35 INFO mapreduce.Job: Job job_1728241920349_1747 running in uber mode : false
24/10/19 10:19:35 INFO mapreduce.Job:  map 0% reduce 0%

24/10/19 10:19:45 INFO mapreduce.Job:  map 67% reduce 0%
24/10/19 10:26:53 INFO mapreduce.Job:  map 68% reduce 0%
24/10/19 10:26:56 INFO mapreduce.Job:  map 69% reduce 0%
24/10/19 10:27:00 INFO mapreduce.Job:  map 70% reduce 0%
24/10/19 10:27:03 INFO mapreduce.Job:  map 72% reduce 0%
24/10/19 10:27:06 INFO mapreduce.Job:  map 75% reduce 0%
24/10/19 10:27:09 INFO mapreduce.Job:  map 77% reduce 0%
24/10/19 10:27:12 INFO mapreduce.Job:  map 79% reduce 0%
24/10/19 10:27:15 INFO mapreduce.Job:  map 82% reduce 0%
24/10/19 10:27:18 INFO mapreduce.Job:  map 84% reduce 0%
24/10/19 10:27:21 INFO mapreduce.Job:  map 87% reduce 0%
24/10/19 10:27:24 INFO mapreduce.Job:  map 89% reduce 0%
24/10/19 10:27:27 INFO mapreduce.Job:  map 91% reduce 0%
24/10/19 10:27:30 INFO mapreduce.Job:  map 93% reduce 0%
24/10/19 10:27:33 INFO mapreduce.Job:  map 95% reduce 0%
24/10/19 10:27:34 INFO mapreduce.Job:  map 96% reduce 0%
24/10/19 10:27:36 INFO mapreduce.Job:  map 97% reduce 0%
24/10/19 10:27:39 INFO mapreduce.Job:  map 98% reduce 0%
24/10/19 10:27:42 INFO mapreduce.Job:  map 99% reduce 0%
24/10/19 10:27:44 INFO mapreduce.Job:  map 100% reduce 0%
24/10/19 10:27:45 INFO mapreduce.Job:  map 100% reduce 5%
24/10/19 10:27:48 INFO mapreduce.Job:  map 100% reduce 16%
24/10/19 10:27:51 INFO mapreduce.Job:  map 100% reduce 20%
24/10/19 10:27:54 INFO mapreduce.Job:  map 100% reduce 21%
24/10/19 10:27:57 INFO mapreduce.Job:  map 100% reduce 22%
24/10/19 10:28:00 INFO mapreduce.Job:  map 100% reduce 23%
24/10/19 10:28:03 INFO mapreduce.Job:  map 100% reduce 24%
24/10/19 10:28:06 INFO mapreduce.Job:  map 100% reduce 25%
24/10/19 10:28:09 INFO mapreduce.Job:  map 100% reduce 26%
```



```
24/10/19 10:29:39 INFO mapreduce.Job:  map 100% reduce 89%
24/10/19 10:29:42 INFO mapreduce.Job:  map 100% reduce 90%
24/10/19 10:29:47 INFO mapreduce.Job:  map 100% reduce 97%
24/10/19 10:29:56 INFO mapreduce.Job:  map 100% reduce 98%
24/10/19 10:30:06 INFO mapreduce.Job:  map 100% reduce 99%
24/10/19 10:30:15 INFO mapreduce.Job:  map 100% reduce 100%
24/10/19 10:30:19 INFO mapreduce.Job: Job job_1728241920349_1747 completed successfully
24/10/19 10:30:19 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=13431328232
                FILE: Number of bytes written=20142924530
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=32598794
                HDFS: Number of bytes written=69536192
                HDFS: Number of read operations=36
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=20
        Job Counters
                Launched map tasks=2
                Launched reduce tasks=10
                Data-local map tasks=2
                Total time spent by all maps in occupied slots (ms)=3850848
                Total time spent by all reduces in occupied slots (ms)=3270744
                Total time spent by all map tasks (ms)=962712
                Total time spent by all reduce tasks (ms)=408843
                Total vcore-milliseconds taken by all map tasks=962712
                Total vcore-milliseconds taken by all reduce tasks=408843
                Total megabyte-milliseconds taken by all map tasks=3943268352
                Total megabyte-milliseconds taken by all reduce tasks=3349241856
        Map-Reduce Framework
                Map input records=1768149
                Map output records=154705236
                Map output bytes=6398752802
                Map output materialized bytes=6713434072
                Input split bytes=252
                Combine input records=0
                Combine output records=0
                Reduce input groups=79555438
                Reduce shuffle bytes=6713434072
                Reduce input records=154705236
                Reduce output records=609090
                Spilled Records=464115708
                Shuffled Maps =20
                Failed Shuffles=0
                Merged Map outputs=20
                GC time elapsed (ms)=2678
                CPU time spent (ms)=1797920
                Physical memory (bytes) snapshot=11238723584
                Virtual memory (bytes) snapshot=121924538368
```



```
                Failed Shuffles=0
                Merged Map outputs=20
                GC time elapsed (ms)=2678
                CPU time spent (ms)=1797920
                Physical memory (bytes) snapshot=11238723584
                Virtual memory (bytes) snapshot=121924538368
```

```
                Total committed heap usage (bytes)=10376183808
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=32598542
        File Output Format Counters
                Bytes Written=69536192
24/10/19 10:30:19 INFO streaming.StreamJob: Output directory: /user/s1155218605/b-output
[s1155218605@dicvmc4 ~]$
[s1155218605@dicvmc4 ~]$
```

D,

Run part (a) for the medium dataset multiple times while modifying the number of mappers and reducers for your MapReduce job(s) each time. You need to examine and report the performance of your program for at least 4 different runs. Each run should use a different combination of the number of mappers and reducers. For each run, performance statistics to be reported should include: (i) the time consumed by the entire MapReduce job(s); (ii) the maximum, minimum and average time consumed by mapper and reducer tasks; (iii) tabulate the time consumption for each MapReduce job and its tasks. (One example is given in the following table.) Moreover, describe (and explain, if possible) your observations.

Job	Mapper num	Reducer num	Max mapper time	Min mapper time	Avg mapper time	Max reducer time	Min reducer time	Avg reducer time	Total time
1	2	4	1min 52s	1min 51s	1min 52s	1min 36s	1min 32s	1min 26s	5min 2
2	4	6	2min 8s	2min 5s	2min 6s	1min 3s	58s	54s	4min 15s
3	8	4	1min 51s	1min 51s	1min 51s	1min 36s	1min 32s	1min 26s	5min 3s
4	2	8	2min 4s	2min 1s	2min 3s	53 s	44s	41s	4min 23s

My observations:

As the number of mappers increases, changes in the number of reducers have an impact on the total job time. In the experiment, Job 1 (with 2 mappers and 4 reducers) and Job 4 (with 2 mappers and 8 reducers) have the same number of mappers, but the difference in the number of reducers leads to a significant difference in total time. Job 4 took 4 minutes and 23 seconds, which is shorter than Job 1's total time of 5 minutes and 2 seconds. This shows that increasing the number of reducers can improve efficiency in some cases.

Part 2

a.

1) Which default ports do machines (VMs) in a multi-node Hadoop cluster use for inter-machine communications (i.e., transmission of network traffic between machines in the cluster)? Name at least 2 ports and describe their roles.

Default ports for communication between machines in a multi-node Hadoop cluster:
Port 50010: Used for transferring data blocks between DataNodes. When one DataNode sends a block to another, this port is used.
Port 8020: The NameNode IPC port. Clients communicate with the NameNode through this port to get file system metadata, like the location of data blocks.

2) Are you using public or private IPs of the VMs to access SSH? Are machines in your Hadoop cluster using public or private IPs to identify and communicate with each other?

I am directly using one of the virtual machines on Google Cloud as the master node and accessing it via the web-based SSH, which utilizes the public IP. However, within the Hadoop cluster, the virtual machines communicate with each other using private IPs.

3) To ensure proper communication between machines, did you set up extra firewall rules/policies as you did for SSH (port 22) in HW#0? Why or why not?

Yes, I also opened TCP, UDP, and ICMP for the private IP ranges 10.0.0.0/8, 172.16.0.0/12, and 192.168.0.0/16.
This ensures smooth communication between machines in the cluster. By allowing all traffic within these common private network ranges, I make sure that nodes can exchange data without being blocked by firewall rules. Since Hadoop uses multiple protocols (TCP, UDP) for different services, opening these ports simplifies configuration and avoids the need to manually open each specific service port. This setup also limits access to internal networks only, enhancing security while maintaining easy communication within the cluster.

b.

Consider a Hadoop cluster with the following configurations:

1, You have allocated 100GB of disk space for each VM in your Google Cloud/AWS Console.
2, There are at most 4 such VMs that can be utilized by the Hadoop cluster. Given this setup, please evaluate the feasibility of taking up 150GB of total disk space on the HDFS (Hadoop Distributed File System). Please list your considerations point by point. You may first give

a general answer and then take into account all potential factors that might affect the usage of disk space.

If using the default replication factor of 3, the cluster won't be able to store 150GB of data. Because HDFS creates 3 copies of each data block by default, meaning each file takes up 3 times the storage space. So, storing 150GB of data actually need 450GB of total storage space.