```
ma4759@login-2-1:~/Bigdata_Final_Project/transportation/etl_code
                                                                                                                                                                                               o. ≡
  [ma4759@login-2-1 etl_code]$ ./etl.sh
  Found 1 items
 -rw-r--r--+ 3 ma4759 users
tialCleaning/Table_data-2.csv
                                                                    426644 2020-12-02 22:58 /user/ma4759/infrastructure/transportation/ini
 Note: Clean.java uses or overrides a deprecated API.
 Note: Recompile with -Xlint:deprecation for details.
 added manifest
 adding: Clean.class(in = 1321) (out= 776)(deflated 41%)
adding: CleanMapper.class(in = 3079) (out= 1310)(deflated 57%)
adding: CleanReducer.class(in = 1145) (out= 416)(deflated 63%)
 20/12/02 22:58:27 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed
. Implement the Tool interface and execute your application with ToolRunner to remedy this.
 20/12/02 22:58:27 INFO input.FileInputFormat: Total input paths to process : 1
20/12/02 22:58:28 INFO mapreduce.JobSubmitter: number of splits:1
 20/12/02 22:58:28 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1604346392376_7799
20/12/02 22:58:28 INFO impl.YarnClientImpl: Submitted application application_1604346392376_7799
20/12/02 22:58:28 INFO mapreduce.Job: The url to track the job: http://babar.es.its.nyu.edu:8088/proxy
  /application 1604346392376 7799/
 20/12/02 22:58:28 INFO mapreduce.Job: Running job: job_1604346392376_7799
20/12/02 22:58:32 INFO mapreduce.Job: Job job_1604346392376_7799 running in uber mode : false
 20/12/02 22:58:32 INFO mapreduce.Job: map 0% reduce 0% 20/12/02 22:58:37 INFO mapreduce.Job: map 100% reduce 0%
 20/12/02 22:58:40 INFO mapreduce.Job: Job job_1604346392376_7799 completed successfully
 20/12/02 22:58:40 INFO mapreduce.Job: Counters: 30
                  File System Counters
                                   FILE: Number of bytes read=0
                                   FILE: Number of bytes written=153912
                                   FILE: Number of read operations=0
FILE: Number of large read operations=0
                                   FILE: Number of write operations=0
                                   HDFS: Number of bytes read=426796
                                   HDFS: Number of bytes written=202493
                                   HDFS: Number of read operations=5
HDFS: Number of large read operations=0
                                   HDFS: Number of write operations=2
                  Job Counters
                                   Launched map tasks=1
                                   Data-local map tasks=1
                                   Total time spent by all maps in occupied slots (ms)=9088
                                   Total time spent by all reduces in occupied slots (ms)=0
                                   Total time spent by all map tasks (ms)=2272
                                   Total vcore-milliseconds taken by all map tasks=2272
                                   Total megabyte-milliseconds taken by all map tasks=9306112
                  Map-Reduce Framework
                                   Map input records=3143
                                   Map output records=3142
                                   Input split bytes=152
                                   Spilled Records=0
                                   Failed Shuffles=0
                                   Merged Map outputs=0
                                   GC time elapsed (ms)=42
CPU time spent (ms)=1370
                                   Physical memory (bytes) snapshot=365891584
Virtual memory (bytes) snapshot=3731124224
Total committed heap usage (bytes)=1194852352
                  File Input Format Counters
                                   Bytes Read=426644
                  File Output Format Counters
                                   Bytes Written=202493
 Alabama, Autauga County, 114, 55050, 0.964912281, 0.035087719, 0, 50.7, 0
 Alabama,Baldwin County,360,199510,0.980555556,0.019444444,0,30.4,0
 Alabama,Barbour County,193,26615,0.968911917,0.031088083,0,14.7,0
 Alabama,Bibb County,126,22570,1,0,0,25.1,5.5
Alabama,Blount County,183,57705,0.912568306,0.087431694,0,13.1,0
Alabama, Bibb County, 126, 22570, 1, 0, 0, 25.1, 5.5
Alabama, Blount County, 183, 57705, 0.912568306, 0.087431694, 0, 13.1, 0
Alabama, Butler County, 138, 10550, 0.985507246, 0.014492754, 0, 0, 0
Alabama, Butler County, 188, 20280, 0.946808511, 0.053191489, 0, 45.3, 0
Alabama, Calhoun County, 374, 115885, 0.962566845, 0.037433155, 0, 65.9, 29.3
Alabama, Chambers County, 184, 34020, 0.929347826, 0.070652174, 0, 28.9, 0
Alabama, Cherokee County, 169, 25895, 0.982248521, 0.017751479, 0, 0, 0
Alabama, Chilton County, 189, 43815, 0.984126984, 0.015873016, 0, 59.4, 0
Alabama, Choctaw County, 153, 13285, 0.934640523, 0.065359477, 0, 26.7, 0
Alabama, Clarke County, 155, 24845, 0.993548387, 0.006451613, 0, 42.9, 0
Alabama, Clay County, 130, 13485, 0.915384615, 0.084615385, 0, 28.8, 0
Alabama, Colbert County, 165, 14990, 0.89090991, 0.109090990, 0, 19.7, 19.7
Alabama, Colbert County, 175, 54375, 0.988571429, 0.011428571, 1, 70.3, 0
Alabama, Conccuh County, 220, 12695, 0.915929204, 0.084070796, 0, 39.1, 0
Alabama, Conoccuh County, 337, 37730, 0.96735905, 0.03264095, 0, 29.2, 0
Alabama, Corington County, 337, 37730, 0.96735905, 0.03264095, 0, 29.2, 0
Alabama, Collman County, 245, 81315, 0.946938776, 0.053061224, 0, 24.6, 0
Alabama, Dale County, 203, 49605, 1, 0, 1, 53.4, 0
Alabama, Dale County, 277, 41425, 0.969162996, 0.030837004, 0, 110.7, 0
Alabama, Dallas County, 277, 41425, 0.969162996, 0.030837004, 0, 110.7, 0
Alabama, Elmore County, 198, 80955, 0.96969697, 0.030837004, 0, 110.7, 0
Alabama, Elmore County, 198, 80955, 0.96969697, 0.03030303, 0, 16.2, 0
Alabama, Escambia County, 196, 37875, 0.964285714.0.035714286.0.89.6.0
```

```
a
                                                      ma4759@login-2-1:~/Bigdata_Final_Project/transportation/etl_code
Note: AcceptableRoads.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
 added manifest
adding: AcceptableRoads.class(in = 1572) (out= 815)(deflated 48%)
adding: AcceptableRoads.class(in = 15/2) (out= 813)(dertated 40%)
adding: AcceptableRoadsMapper1.class(in = 1753) (out= 705)(deflated 59%)
adding: AcceptableRoadsMapper2.class(in = 1812) (out= 726)(deflated 59%)
adding: AcceptableRoadsReducer.class(in = 1999) (out= 872)(deflated 56%)
20/12/02 22:58:57 WARN mapreduce.JobResourceUploader: Haddoop command-line option parsing not performed
20/12/02 22:58:58 INFO input.FileInputFormat: Total input paths to process: 1
20/12/02 22:58:58 INFO input.FileInputFormat: Total input paths to process: 1
20/12/02 22:58:58 INFO input.FileInputFormat: Total input paths to process: 1
20/12/02 22:58:58 INFO mapreduce.JobSubmitter: number of splits:2
20/12/02 22:58:59 INFO mapreduce.JobSubmitter: Number of sptis.2
20/12/02 22:58:59 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1604346392376_7800
20/12/02 22:58:59 INFO impl.YarnClientImpl: Submitted application application_1604346392376_7800
20/12/02 22:58:59 INFO mapreduce.Job: The url to track the job: http://babar.es.its.nyu.edu:8088/proxy
/application_1604346392376_7800/
20/12/02 22:58:59 INFO mapreduce.Job: Running job: job_1604346392376_7800
20/12/02 22:59:03 INFO mapreduce.Job: Job job_1604346392376_7800 running in uber mode : false
20/12/02 22:59:03 INFO mapreduce.Job: map 0% reduce 0%
20/12/02 22:59:07 INFO mapreduce.Job: map 100% reduce 0%
20/12/02 22:59:13 INFO mapreduce.Job: map 100% reduce 100%
20/12/02 22:59:19 INFO mapreduce.Job: Job job_1604346392376_7800 completed successfully
20/12/02 22:59:19 INFO mapreduce.Job: Counters: 49
              File System Counters
                            FILE: Number of bytes read=137788
FILE: Number of bytes written=5542767
FILE: Number of read operations=0
FILE: Number of large read operations=0
                            FILE: Number of write operations=0
HDFS: Number of bytes read=204215
HDFS: Number of bytes written=239973
                             HDFS: Number of read operations=102
                             HDFS: Number of large read operations=0
                             HDFS: Number of write operations=64
              Job Counters
                             Launched map tasks=2
                             Launched reduce tasks=32
                             Data-local map tasks=2
                             Total time spent by all maps in occupied slots (ms)=17432
                             Total time spent by all reduces in occupied slots (ms)=456294
                             Total time spent by all map tasks (ms)=4358
                             Total time spent by all reduce tasks (ms)=76049
                             Total vcore-milliseconds taken by all map tasks=4358
                            Total vcore-milliseconds taken by all reduce tasks=76049
Total megabyte-milliseconds taken by all map tasks=17850368
Total megabyte-milliseconds taken by all reduce tasks=467245056
              Map-Reduce Framework
                             Map input records=3195
                             Map output records=3194
                             Map output bytes=231686
                             Map output materialized bytes=138527
                            Input split bytes=565
                            Combine input records=0
                             Combine output records=0
                            Reduce input groups=52
Reduce shuffle bytes=138527
                             Reduce input records=3194
                             Reduce output records=3142
                             Spilled Records=6388
                            Shuffled Maps =64
                            Failed Shuffles=0
                            Merged Map outputs=64
GC time elapsed (ms)=1673
                            GC time etapsed (ms)=42810
CPU time spent (ms)=42810
Physical memory (bytes) snapshot=13691826176
Virtual memory (bytes) snapshot=127226134528
Total committed heap usage (bytes)=40938504192
              Shuffle Errors
                             BAD_ID=0
                             CONNECTION=0
                             TO FRROR=0
                            WRONG_LENGTH=0
                            WRONG_MAP=0
                            WRONG_REDUCE=0
              File Input Format Counters
                            Bytes Read=0
              File Output Format Counters
                            Bytes Written=239973
 Nevada,Washoe County,299,439915,0.97993311,0.02006689,1,143.5,42,0.873795386
Nevada, White Pine County,5,9895,1,0,0,31.9,31.9,0.873795386
Nevada,Carson City,0,54410,0,0,0,0,0,0.873795386
Nevada,Churchill County,68,24150,0.985294118,0.014705882,0,68.5,42,0.873795386
Nevada,Clark County,1056,2070155,0.996212121,0.003787879,1,160.2,4.3,0.873795386
```

```
ma4759@login-2-1:~/Bigdata_Final_Project/transportation/etl_code
adding: LandAreaReducer.class(in = 2254) (out= 1016)(deflated 54%)
20/12/02 22:59:37 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed
. Implement the Tool interface and execute your application with ToolRunner to remedy this.
20/12/02 22:59:38 INFO input.FileInputFormat: Total input paths to process : 1
20/12/02 22:59:38 INFO input.FileInputFormat: Total input paths to process : 1
20/12/02 22:59:38 INFO mapreduce.JobSubmitter: number of splits:2
20/12/02 22:59:38 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1604346392376_7801
20/12/02 22:59:39 INFO impl.YarnClientImpl: Submitted application application_1604346392376_7801
20/12/02 22:59:39 INFO mapreduce.Job: The url to track the job: http://babar.es.its.nyu.edu:8088/proxy
/application_1604346392376_7801/
7app:rcarion_10043639270_78017
20/12/02 22:59:39 INFO mapreduce.Job: Running job: job_1604346392376_7801
20/12/02 22:59:43 INFO mapreduce.Job: Job job_1604346392376_7801 running in uber mode : false
20/12/02 22:59:43 INFO mapreduce.Job: map 0% reduce 0%
20/12/02 22:59:47 INFO mapreduce.Job: map 100% reduce 0%
20/12/02 22:59:53 INFO mapreduce.Job: map 100% reduce 100%
20/12/02 22:59:57 INFO mapreduce.Job: Job job_1604346392376_7801 completed successfully
20/12/02 22:59:57 INFO mapreduce.Job: Counters: 50
            File System Counters
                         FILE: Number of bytes read=259334
                         FILE: Number of bytes written=5815453
                         FILE: Number of read operations=0
                         FILE: Number of large read operations=0
FILE: Number of write operations=0
                         HDFS: Number of bytes read=785943
                         HDFS: Number of bytes written=223243
                         HDFS: Number of read operations=102
                         HDFS: Number of large read operations=0
                         HDFS: Number of write operations=64
             Job Counters
                         Launched map tasks=2
                         Launched reduce tasks=32
                         Data-local map tasks=1
                         Rack-local map tasks=1
                         Total time spent by all maps in occupied slots (ms)=18896
                         Total time spent by all reduces in occupied slots (ms)=466728
Total time spent by all map tasks (ms)=4724
                         Total time spent by all reduce tasks (ms)=77788
Total vcore-milliseconds taken by all map tasks=4724
                         Total vcore-milliseconds taken by all reduce tasks=77788
                          Total megabyte-milliseconds taken by all map tasks=19349504
                         Total megabyte-milliseconds taken by all reduce tasks=477929472
            Map-Reduce Framework
                         Map input records=6341
                         Map output records=6287
                         Map output bytes=401153
                         Map output materialized bytes=290687
                         Input split bytes=550
                         Combine input records=0
                         Combine output records=0
                         Reduce input groups=3147
                         Reduce shuffle bytes=290687
                         Reduce input records=6287
                         Reduce output records=3129
                         Spilled Records=12574
                         Shuffled Maps =64
Failed Shuffles=0
                         Merged Map outputs=64
                         GC time elapsed (ms)=1938
                         CPU time spent (ms)=49520
                         Physical memory (bytes) snapshot=14309101568
Virtual memory (bytes) snapshot=127259836416
                          Total committed heap usage (bytes)=41350070272
             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=0
             File Output Format Counters
                         Bytes Written=223243
Bytes Written=223243
Michigan,Alpena,32,28930,0.9375,0.0625,19,0.771767308,1695
Mississippi,Amite,230,12690,0.804347826,0.195652174,0,0.731516457,731.64
North Carolina,Ashe,242,26990,0.723140496,0.276859504,0,0.890560355,426.78
Virginia,Bedford,196,76935,0.887755102,0.112244898,84.7,0.884284057,6.9
Kentucky,Bell,206,27675,0.878640777,0.121359223,78,0.920222177,361.35
South Carolina,Berkeley,206,199080,0.941747573,0.058252427,113.7,0.812866638,1228.07
Pennsylvania,Berks,665,414095,0.784962406,0.215037594,121.6,0.714875343,865.67
Wyoming Rig Horn 11330 0 001008201 0 008801009 75 9 0 803817764 3158 96
Wyoming,Big Horn,11930,0.901098901,0.098901099,75.9,0.803817764,3158.96
Georgia, Bleckley, 12835, 0.965517241, 0.034482759, 16, 0.967672415, 219.14
Idaho, Boise, 88, 6890, 0.943181818, 0.056818182, 31.9, 0.968883613, 1906.76
```