

1. `javac -classpath "$(yarn classpath)" -d . CleanMapper.java`

```
[lam923@login-2-1 ~]$ javac -classpath "$(yarn classpath)" -d . CleanMapper.java
[lam923@login-2-1 ~]$
```

2. `javac -classpath "$(yarn classpath)" -d . CleanReducer.java`

```
[lam923@login-2-1 ~]$ javac -classpath "$(yarn classpath)" -d . CleanReducer.java
[lam923@login-2-1 ~]$
```

3. `javac -classpath "$(yarn classpath)":.. -d . Clean.java`

```
[lam923@login-2-1 ~]$ javac -classpath "$(yarn classpath)":.. -d . Clean.java
Note: Clean.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
```

4. `jar -cvf Clean.jar *.class`

```
[lam923@login-2-1 ~]$ jar -cvf Clean.jar *.class
added manifest
adding: Clean.class(in = 1190) (out= 691)(deflated 41%)
adding: CleanMapper.class(in = 2721) (out= 1325)(deflated 51%)
adding: CleanReducer.class(in = 1762) (out= 734)(deflated 58%)
adding: CountRecs.class(in = 1257) (out= 719)(deflated 42%)
adding: CountRecsMapper.class(in = 1554) (out= 601)(deflated 61%)
adding: CountRecsReducer.class(in = 1602) (out= 667)(deflated 58%)
adding: CountTotalCases.class(in = 1288) (out= 730)(deflated 43%)
adding: CountTotalCasesMapper.class(in = 1955) (out= 792)(deflated 59%)
adding: CountTotalCasesReducer.class(in = 1614) (out= 672)(deflated 58%)
adding: MaxTemperature.class(in = 1452) (out= 822)(deflated 43%)
adding: MaxTemperatureMapper.class(in = 1876) (out= 805)(deflated 57%)
adding: MaxTemperatureReducer.class(in = 1687) (out= 714)(deflated 57%)
adding: NhbdCount.class(in = 1272) (out= 728)(deflated 42%)
adding: NhbdCountMapper.class(in = 2275) (out= 965)(deflated 57%)
adding: NhbdCountReducer.class(in = 1602) (out= 667)(deflated 58%)
adding: OrganizeMonths.class(in = 1282) (out= 727)(deflated 43%)
adding: OrganizeMonthsMapper.class(in = 2010) (out= 806)(deflated 59%)
adding: OrganizeMonthsReducer.class(in = 1851) (out= 787)(deflated 57%)
adding: WordCount.class(in = 1257) (out= 722)(deflated 42%)
adding: WordCountMapper.class(in = 1927) (out= 795)(deflated 58%)
adding: WordCountReducer.class(in = 1602) (out= 667)(deflated 58%)
```

5. `hdfs dfs -mkdir CleanOutput`

```
[lam923@login-2-1 ~]$ hdfs dfs -mkdir CleanOutput
[lam923@login-2-1 ~]$
```

6. `hdfs dfs -put covid_data.csv CleanOutput`

```
[lam923@login-2-1 ~]$ hdfs dfs -put covid_data.csv CleanOutput
[lam923@login-2-1 ~]$
```

7. `hadoop jar Clean.jar Clean /user/lam923/CleanOutput/covid_data.csv /user/lam923/CleanOutput/clean_output`
(this screenshot shows `hadoop jar Clean.jar Clean /user/lam923/CleanOutput/covid_data.csv /user/lam923/CleanOutput/clean_output2`. The same command for `/user/lam923/CleanOutput/covid_data.csv /user/lam923/CleanOutput/clean_output` was used, but the screenshot quality turned out poor for that command, so I just ran it again using `clean_output2`. The output of both `clean_output` and `clean_output2` are shown below in the next step(they have the same output))

```
[lam923@login-1-1 ~]$ hadoop jar Clean.jar Clean /user/lam923/CleanOutput/covid_data.csv /user/lam923/CleanOutput/clean_output2
20/12/02 09:04:39 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 09:04:40 INFO input.FileInputFormat: Total input paths to process : 1
20/12/02 09:04:41 INFO mapreduce.JobSubmitter: number of splits:1
20/12/02 09:04:41 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1604346392376_7201
20/12/02 09:04:41 INFO impl.YarnClientImpl: Submitted application application_1604346392376_7201
20/12/02 09:04:41 INFO mapreduce.Job: The url to track the job: http://babar.es.its.nyu.edu:8088/proxy/application_1604346392376_7201/
20/12/02 09:04:41 INFO mapreduce.Job: Running job: job_1604346392376_7201
20/12/02 09:04:45 INFO mapreduce.Job: Job job_1604346392376_7201 running in uber mode : false
20/12/02 09:04:45 INFO mapreduce.Job: map 0% reduce 0%
20/12/02 09:04:50 INFO mapreduce.Job: map 100% reduce 0%
20/12/02 09:04:54 INFO mapreduce.Job: map 100% reduce 100%
20/12/02 09:04:55 INFO mapreduce.Job: Job job_1604346392376_7201 completed successfully
20/12/02 09:04:55 INFO mapreduce.Job: Counters: 49
File System Counters
FILE: Number of bytes read=179883
FILE: Number of bytes written=668001
```

8. `hdfs dfs -cat /user/lam923/CleanOutput/clean_output/part-r-00000`

```
[lam923@login-2-1 ~]$ hdfs dfs -cat /user/lam923/CleanOutput/clean_output/part-r-00000
1/22/20 , AL , 0 , 0
1/22/20 , AR , 0 , 0
1/22/20 , AZ , 0 , 0
1/22/20 , CA , 0 , 0
1/22/20 , CO , 0 , 0
1/22/20 , CT , 0 , 0
1/22/20 , DE , 0 , 0
1/22/20 , FL , 0 , 0
1/22/20 , GA , 0 , 0
1/22/20 , HI , 0 , 0
1/22/20 , IA , 0 , 0
1/22/20 , ID , 0 , 0
1/22/20 , IL , 0 , 0
1/22/20 , IN , 0 , 0
1/22/20 , KS , 0 , 0
1/22/20 , KY , 0 , 0
```

```

9/9/20 , NY , 204307 , 303
9/9/20 , OH , 132965 , 973
9/9/20 , OK , 71877 , 693
9/9/20 , OR , 28471 , 116
9/9/20 , PA , 141290 , 931
9/9/20 , RI , 22676 , 84
9/9/20 , SC , 126213 , 305
9/9/20 , SD , 15571 , 168
9/9/20 , TN , 166587 , 833
9/9/20 , TX , 645791 , 4000
9/9/20 , UT , 56401 , 383
9/9/20 , VA , 130529 , 1240
9/9/20 , VT , 1661 , 7
9/9/20 , WA , 78009 , 464
9/9/20 , WI , 88590 , 913
9/9/20 , WV , 11808 , 147
9/9/20 , WY , 4151 , 48
[lam923@login-1-1 ~]$

```

hdfs dfs -cat /user/lam923/CleanOutput/clean_output2/part-r-00000

```

[lam923@login-1-1 ~]$ hdfs dfs -cat /user/lam923/CleanOutput/clean_output2/part-r-00000
1/22/20 , AL , 0 , 0
1/22/20 , AR , 0 , 0
1/22/20 , AZ , 0 , 0
1/22/20 , CA , 0 , 0
1/22/20 , CO , 0 , 0
1/22/20 , CT , 0 , 0
1/22/20 , DE , 0 , 0
1/22/20 , FL , 0 , 0
1/22/20 , GA , 0 , 0
1/22/20 , HI , 0 , 0
1/22/20 , IA , 0 , 0
1/22/20 , ID , 0 , 0

```

9/9/20	, NY ,	204307 ,	303
9/9/20	, OH ,	132965 ,	973
9/9/20	, OK ,	71877 ,	693
9/9/20	, OR ,	28471 ,	116
9/9/20	, PA ,	141290 ,	931
9/9/20	, RI ,	22676 ,	84
9/9/20	, SC ,	126213 ,	305
9/9/20	, SD ,	15571 ,	168
9/9/20	, TN ,	166587 ,	833
9/9/20	, TX ,	645791 ,	4000
9/9/20	, UT ,	56401 ,	383
9/9/20	, VA ,	130529 ,	1240
9/9/20	, VT ,	1661 ,	7
9/9/20	, WA ,	78009 ,	464
9/9/20	, WI ,	88590 ,	913
9/9/20	, WV ,	11808 ,	147
9/9/20	, WY ,	4151 ,	48