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

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

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

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

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

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

4. jar -cvf CountRecs.jar *.class

```
[lam923@login-1-1 ~]$  jar -cvf CountRecs.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 = 1398) (out= 532)(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 CountRecs

```
user/lam923/input_data
[lam923@login-1-1 ~]$ hdfs dfs -mkdir CountRecs
```

6. hdfs dfs -put covid_data.csv CountRecs

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

7. hdfs dfs -get /user/lam923/CleanOutput/clean_output

```
[lam923@login-2-1 ~]$ hdfs dfs -get /user/lam923/CleanOutput/clean_output
```

8. hdfs dfs -put clean_output CountRecs

```
[lam923@login-2-1 ~]$ hdfs dfs -put clean_output CountRecs
```

9. hadoop jar CountRecs.jar CountRecs /user/lam923/CountRecs/covid_data.csv /user/lam923/CountRecs/original_count

```
[lam923@login-2-1 ~]$ hadoop jar CountRecs.jar CountRecs /user/lam923/CountRecs/covid_data.csv /user/lam9
3/CountRecs/original_count
20/12/02 00:42:32 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. I
plement the Tool interface and execute your application with ToolRunner to remedy this.
20/12/02 00:42:32 INFO input.FileInputFormat: Total input paths to process : 1
20/12/02 00:42:34 INFO mapreduce.JobSubmitter: number of splits:1
20/12/02 00:42:55 INFO hdfs.DFSClient: Exception in createBlockOutputStream
java.io.IOException: Bad connect ack with firstBadLink as 10.0.255.225:50010
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.createBlockOutputStream(DFSOutputStream.ja
a:1785)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.nextBlockOutputStream(DFSOutputStream.java
1683)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:790)
20/12/02 00:42:55 WARN hdfs.DFSClient: Abandoning BP-1256044058-128.122.215.50-1440607644284:blk_11673562
7_93834311
20/12/02 00:42:55 WARN hdfs.DFSClient: Excluding datanode DatanodeInfoWithStorage[10.0.255.225:50010,DS-9
b24b2c-0444-47ac-af35-546b9a827a69,DISK]
20/12/02 00:42:55 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1604346392376_7066
20/12/02 00:42:56 INFO impl.YarnClientImpl: Submitted application application_1604346392376_7066
20/12/02 00:42:56 INFO mapreduce.Job: The url to track the job: http://babar.es.its.nyu.edu:8088/proxy/ap
```

10. hdfs dfs -cat /user/lam923/CountRecs/original_count/part-r-00000

```
[[lam923@login-2-1 ~]$ hdfs dfs -cat /user/lam923/CountRecs/original_count/part-r-00000
 count   17401
```

11. hadoop jar CountRecs.jar CountRecs /user/lam923/CountRecs/clean_output/part-r-00000 /user/lam923/CountRecs/new_count

```
[lam923@login-2-1 ~]$ hadoop jar CountRecs.jar CountRecs /user/lam923/CountRecs/covid_data.csv /user/lam92
3/CountRecs/original_count
20/12/02 00:42:32 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Im
plement the Tool interface and execute your application with ToolRunner to remedy this.
20/12/02 00:42:32 INFO input.FileInputFormat: Total input paths to process : 1
20/12/02 00:42:34 INFO mapreduce.JobSubmitter: number of splits:1
20/12/02 00:42:55 INFO hdfs.DFSClient: Exception in createBlockOutputStream
java.io.IOException: Bad connect ack with firstBadLink as 10.0.255.225:50010
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.createBlockOutputStream(DFSOutputStream.jav
a:1785)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.nextBlockOutputStream(DFSOutputStream.java:
1683)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:790)
20/12/02 00:42:55 WARN hdfs.DFSClient: Abandoning BP-1256044058-128.122.215.50-1440607644284:blk_116735622
7_93834311
20/12/02 00:42:55 WARN hdfs.DFSClient: Excluding datanode DatanodeInfoWithStorage[10.0.255.225:50010,DS-97
b24b2c-0444-47ac-af35-546b9a827a69,DISK]
20/12/02 00:42:55 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1604346392376_7066
20/12/02 00:42:56 INFO impl.YarnClientImpl: Submitted application application_1604346392376_7066
20/12/02 00:42:56 INFO mapreduce.Job: The url to track the job: http://babar.es.its.nyu.edu:8088/proxy/app
lication_1604346392376_7066/
20/12/02 00:42:56 INFO mapreduce.Job: Running job: job_1604346392376_7066
20/12/02 00:43:00 INFO mapreduce.Job: Job job_1604346392376_7066 running in uber mode : false
20/12/02 00:43:00 INFO mapreduce.Job:  map 0% reduce 0%
20/12/02 00:43:05 INFO mapreduce.Job:  map 100% reduce 0%
20/12/02 00:43:09 INFO mapreduce.Job:  map 100% reduce 100%
20/12/02 00:43:26 INFO mapreduce.Job: Job job_1604346392376_7066 completed successfully
20/12/02 00:43:26 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=9902
```

12. hdfs dfs -cat /user/lam923/CountRecs/new_count/part-r-00000

```
[[lam923@login-2-1 ~]$ hdfs dfs -cat /user/lam923/CountRecs/new_count/part-r-00000
 count   14144
```