```
bfezo001@lena:~$ # removes existing output
bfezo001@lena:~$ hadoop fs -rm -r cwq1_output
Deleted cwq1_output
bfezo001@lena:~$ hadoop fs -rm -r dsm010/cw_question_1
Deleted dsm010/cw_question_1
bfezo001@lena:~$ hadoop fs -copyFromLocal dsm010/cw_question_1 dsm010
bfezo001@lena:~$ # executes the job
bfezo001@lena:~$ hadoop jar /opt/hadoop/current/share/hadoop/tools/lib/hadoop-
streaming-3.3.0.jar \
>      -input dsm010/cw_question_1/200707hourly.txt \
>      -output cwq1_output \
>      -file dsm010/cw_question_1/mapper.py \
>      -mapper mapper.py \
>      -file dsm010/cw_question_1/reducer.py \
>      -reducer reducer.py
2023-01-04 13:14:22,824 WARN streaming.StreamJob: -file option is deprecated,
please use generic option -files instead.
packageJobJar: [dsm010/cw_question_1/mapper.py, dsm010/cw_question_1/reducer.py,
/tmp/hadoop-unjar12797987160986315593/] [] /tmp/streamjob17998758898595020770.jar
tmpDir=null
2023-01-04 13:14:23,396 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting
to ResourceManager at lena-master/128.86.245.64:8032
2023-01-04 13:14:23,488 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting
to ResourceManager at lena-master/128.86.245.64:8032
2023-01-04 13:14:23,672 INFO mapreduce.JobResourceUploader: Disabling Erasure
Coding for path: /tmp/hadoop-yarn/staging/bfezo001/.staging/job_1672817468245_0031
2023-01-04 13:14:23,979 INFO mapred.FileInputFormat: Total input files to process :
1
2023-01-04 13:14:24,082 INFO mapreduce.JobSubmitter: number of splits:2
2023-01-04 13:14:24,221 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1672817468245_0031
2023-01-04 13:14:24,222 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-01-04 13:14:24,332 INFO conf.Configuration: resource-types.xml not found
2023-01-04 13:14:24,332 INFO resource.ResourceUtils: Unable to find 'resource-
types.xml'.
2023-01-04 13:14:24,587 INFO impl.YarnClientImpl: Submitted application
application_1672817468245_0031
2023-01-04 13:14:24,631 INFO mapreduce.Job: The url to track the job: http://lena-
master:8088/proxy/application_1672817468245_0031/
2023-01-04 13:14:24,633 INFO mapreduce.Job: Running job: job_1672817468245_0031
2023-01-04 13:14:29,720 INFO mapreduce.Job: Job job_1672817468245_0031 running in
uber mode : false
2023-01-04 13:14:29,721 INFO mapreduce.Job:  map 0% reduce 0%
2023-01-04 13:14:35,808 INFO mapreduce.Job:  map 100% reduce 0%
2023-01-04 13:14:39,844 INFO mapreduce.Job:  map 100% reduce 100%
2023-01-04 13:14:40,864 INFO mapreduce.Job: Job job_1672817468245_0031 completed
successfully
2023-01-04 13:14:40,953 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=14023353
                FILE: Number of bytes written=28852301
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=101449659
                HDFS: Number of bytes written=729
                HDFS: Number of read operations=11
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
```

```
                        HDFS: Number of bytes read erasure-coded=0
            Job Counters
                        Launched map tasks=2
                        Launched reduce tasks=1
                        Rack-local map tasks=2
                        Total time spent by all maps in occupied slots (ms)=27880
                        Total time spent by all reduces in occupied slots (ms)=13740
                        Total time spent by all map tasks (ms)=5576
                        Total time spent by all reduce tasks (ms)=2748
                        Total vcore-milliseconds taken by all map tasks=5576
                        Total vcore-milliseconds taken by all reduce tasks=2748
                        Total megabyte-milliseconds taken by all map tasks=28549120
                        Total megabyte-milliseconds taken by all reduce tasks=14069760
            Map-Reduce Framework
                        Map input records=785807
                        Map output records=605083
                        Map output bytes=12813181
                        Map output materialized bytes=14023359
                        Input split bytes=244
                        Combine input records=0
                        Combine output records=0
                        Reduce input groups=222738
                        Reduce shuffle bytes=14023359
                        Reduce input records=605083
                        Reduce output records=20
                        Spilled Records=1210166
                        Shuffled Maps =2
                        Failed Shuffles=0
                        Merged Map outputs=2
                        GC time elapsed (ms)=73
                        CPU time spent (ms)=7350
                        Physical memory (bytes) snapshot=1047105536
                        Virtual memory (bytes) snapshot=19129757696
                        Total committed heap usage (bytes)=3158310912
                        Peak Map Physical memory (bytes)=376762368
                        Peak Map Virtual memory (bytes)=6373556224
                        Peak Reduce Physical memory (bytes)=299511808
                        Peak Reduce Virtual memory (bytes)=6389809152
            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=101449415
            File Output Format Counters
                        Bytes Written=729
2023-01-04 13:14:40,953 INFO streaming.StreamJob: Output directory: cwq1_output
bfezo001@lena:~$
bfezo001@lena:~$ # shows the output
bfezo001@lena:~$ hadoop fs -cat cwq1_output/part-00000
20070701, 31, 2, 54.4, 179.319630
20070702, 80, 2, 54.9, 172.632865
20070703, 37, 3, 57.3, 157.355049
20070704, 29, 3, 59.6, 123.422502
20070705, 37, 3, 60.2, 108.453107
20070706, 64, 3, 59.4, 105.319097
```

```
20070707, 51, 3, 59.8, 116.194016
20070708, 37, 3, 61.2, 121.792187
20070709, 36, 2, 61.5, 132.658168
20070710, 40, 4, 61.2, 137.674701
20070711, 37, 2, 59.1, 125.504252
20070712, 32, 4, 56.9, 104.966661
20070713, 31, 3, 57.0, 102.279520
20070714, 30, 5, 58.2, 92.490864
20070715, 36, 6, 59.1, 98.703993
20070716, 35, 6, 59.6, 103.332373
20070717, 35, 4, 61.6, 93.915360
20070718, 39, 4, 63.6, 87.632828
20070719, 12, 63, 74.8, 2.007822
[[1.0, -0.426, 0.373], [-0.426, 1.0, -0.942], [0.373, -0.942, 1.0]]
```