

1. Copy of Your mapper.py code (or equivalent in another programming language) (25% of total grade) – see github mapper.py
2. Copy of Your reducer.py code (or equivalent in another programming language) (25% of total grade) – see github reducer.py
3. Screenshot of the execution of Hadoop MapReduce Job in the terminal (25% of total grade)

```

zihaos_gcp@cluster-12dc-m:~/mapreduce$ hadoop jar /usr/lib/hadoop/hadoop-streaming.jar -file mapper.py -mapper '
python mapper.py' -file reducer.py -reducer 'python reducer.py' -input /data/* -output /OutputFolder
2021-11-18 19:22:58,608 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files i
nstead.
packageJobJar: [mapper.py, reducer.py] [/usr/lib/hadoop/hadoop-streaming-3.2.2.jar] /tmp/streamjob75369520515640
9899.jar tmpDir=null
2021-11-18 19:22:59,463 INFO client.RMPProxy: Connecting to ResourceManager at cluster-12dc-m/10.128.0.3:8032
2021-11-18 19:22:59,717 INFO client.AHSProxy: Connecting to Application History server at cluster-12dc-m/10.128.
0.3:10200
2021-11-18 19:23:00,226 INFO client.RMPProxy: Connecting to ResourceManager at cluster-12dc-m/10.128.0.3:8032
2021-11-18 19:23:00,226 INFO client.AHSProxy: Connecting to Application History server at cluster-12dc-m/10.128.
0.3:10200
2021-11-18 19:23:00,426 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/
staging/zihaos_gcp/.staging/job_1637262641512_0001
2021-11-18 19:23:00,849 INFO mapred.FileInputFormat: Total input files to process : 2
2021-11-18 19:23:00,942 INFO mapreduce.JobSubmitter: number of splits:22
2021-11-18 19:23:01,122 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1637262641512_0001
2021-11-18 19:23:01,125 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-11-18 19:23:01,373 INFO conf.Configuration: resource-types.xml not found
2021-11-18 19:23:01,373 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-11-18 19:23:01,784 INFO impl.YarnClientImpl: Submitted application application_1637262641512_0001
2021-11-18 19:23:01,845 INFO mapreduce.Job: The url to track the job: http://cluster-12dc-m:8088/proxy/applicati
on_1637262641512_0001/
2021-11-18 19:23:01,848 INFO mapreduce.Job: Running job: job_1637262641512_0001
2021-11-18 19:23:13,088 INFO mapreduce.Job: Job job_1637262641512_0001 running in uber mode : false
2021-11-18 19:23:13,089 INFO mapreduce.Job: map 0% reduce 0%
2021-11-18 19:23:22,243 INFO mapreduce.Job: map 14% reduce 0%
2021-11-18 19:23:25,266 INFO mapreduce.Job: map 32% reduce 0%
2021-11-18 19:23:28,283 INFO mapreduce.Job: map 45% reduce 0%
2021-11-18 19:23:31,309 INFO mapreduce.Job: map 50% reduce 0%
2021-11-18 19:23:33,319 INFO mapreduce.Job: map 64% reduce 0%
2021-11-18 19:23:35,336 INFO mapreduce.Job: map 73% reduce 0%
2021-11-18 19:23:36,341 INFO mapreduce.Job: map 77% reduce 0%
2021-11-18 19:23:37,346 INFO mapreduce.Job: map 82% reduce 0%
2021-11-18 19:23:39,356 INFO mapreduce.Job: map 86% reduce 0%
2021-11-18 19:23:40,361 INFO mapreduce.Job: map 100% reduce 0%
2021-11-18 19:23:47,401 INFO mapreduce.Job: map 100% reduce 29%
2021-11-18 19:23:49,411 INFO mapreduce.Job: map 100% reduce 57%
2021-11-18 19:23:50,417 INFO mapreduce.Job: map 100% reduce 86%
2021-11-18 19:23:51,421 INFO mapreduce.Job: map 100% reduce 100%
2021-11-18 19:23:53,441 INFO mapreduce.Job: Job job_1637262641512_0001 completed successfully
2021-11-18 19:23:53,531 INFO mapreduce.Job: Counters: 55
    File System Counters
        FILE: Number of bytes read=192588
        FILE: Number of bytes written=7588799
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=1860914
        HDFS: Number of bytes written=9111
        HDFS: Number of read operations=101
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=21
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Killed reduce tasks=1
        Launched map tasks=22
        Launched reduce tasks=7
        Data-local map tasks=22
        Total time spent by all maps in occupied slots (ms)=448016292

```

4. Copy of your output file (after merging) containing the results (25% of total grade)

See github sorted_output