

Name and Roll No: Malhar Shah (18BCE111)

Date: 14-10-2021

Course Name: Big Data Analytics (2CS702)

Practical - 5: Apply MapReduce algorithms to find phrase frequency from a given dataset.

- Prepare a report to guide design of mapper and reducer.

Output

```
Administrator: Command Prompt
C:\Users\Malhar\Desktop>hadoop fs -mkdir /input_lengthfreq

C:\Users\Malhar\Desktop>hadoop fs -copyFromLocal lfreq.txt /input_lengthfreq
2021-10-14 19:46:32,818 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false

C:\Users\Malhar\Desktop>hadoop jar Lab_5_MapReducePrograms.jar FindPhraseFrequency /input_lengthfreq /output_lengthfreq
2021-10-14 19:47:27,365 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2021-10-14 19:47:28,063 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-10-14 19:47:28,085 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/Malhar/.staging/job_1634207977401_0004
2021-10-14 19:47:28,201 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-10-14 19:47:28,348 INFO input.FileInputFormat: Total input files to process : 1
2021-10-14 19:47:28,398 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-10-14 19:47:28,438 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-10-14 19:47:28,454 INFO mapreduce.JobSubmitter: number of splits:1
2021-10-14 19:47:28,583 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2021-10-14 19:47:28,605 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1634207977401_0004
2021-10-14 19:47:28,606 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-10-14 19:47:28,803 INFO conf.Configuration: resource-types.xml not found
2021-10-14 19:47:28,804 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
```

```
Administrator: Command Prompt
2021-10-14 19:47:28,876 INFO impl.YarnClientImpl: Submitted application application_1634207977401_0004
2021-10-14 19:47:28,922 INFO mapreduce.Job: The url to track the job: http://DESKTOP-LAAH52F:8088/proxy/application_1634207977401_0004/
2021-10-14 19:47:28,922 INFO mapreduce.Job: Running job: job_1634207977401_0004
2021-10-14 19:47:57,477 INFO mapreduce.Job: Job job_1634207977401_0004 running in uber mode : false
2021-10-14 19:47:57,481 INFO mapreduce.Job: map 0% reduce 0%
2021-10-14 19:48:05,067 INFO mapreduce.Job: map 100% reduce 0%
2021-10-14 19:48:12,173 INFO mapreduce.Job: map 100% reduce 100%
2021-10-14 19:48:12,185 INFO mapreduce.Job: Job job_1634207977401_0004 completed successfully
2021-10-14 19:48:12,364 INFO mapreduce.Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=126
    FILE: Number of bytes written=453497
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=172
    HDFS: Number of bytes written=24
    HDFS: Number of read operations=8
    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=1
    Launched reduce tasks=1
    Data-local map tasks=1
```

```
Administrator: Command Prompt

Total time spent by all maps in occupied slots (ms)=4261
Total time spent by all reduces in occupied slots (ms)=5047
Total time spent by all map tasks (ms)=4261
Total time spent by all reduce tasks (ms)=5047
Total vcore-milliseconds taken by all map tasks=4261
Total vcore-milliseconds taken by all reduce tasks=5047
Total megabyte-milliseconds taken by all map tasks=4363264
Total megabyte-milliseconds taken by all reduce tasks=5168128

Map-Reduce Framework
  Map input records=2
  Map output records=12
  Map output bytes=96
  Map output materialized bytes=126
  Input split bytes=113
  Combine input records=0
  Combine output records=0
  Reduce input groups=6
  Reduce shuffle bytes=126
  Reduce input records=12
  Reduce output records=6
  Spilled Records=24
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=122
  CPU time spent (ms)=1653
```

```
Administrator: Command Prompt

Physical memory (bytes) snapshot=505753600
Virtual memory (bytes) snapshot=726269952
Total committed heap usage (bytes)=341835776
Peak Map Physical memory (bytes)=303661056
Peak Map Virtual memory (bytes)=415076352
Peak Reduce Physical memory (bytes)=202092544
Peak Reduce Virtual memory (bytes)=311193600

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=59
File Output Format Counters
  Bytes Written=24
```

```
Administrator: Command Prompt

C:\Users\Malhar\Desktop>hadoop fs -cat /output_lengthfreq/part-r-00000
2021-10-14 19:49:44,617 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2      3
3      3
4      3
5      1
6      1
9      1
```

Conclusion: Running the above map-reduce program we got to practice with the mapper and reducer to perform various other tasks.