EXP NO: 2

Run A Basic Word Count Map Reduce Program To
Understand Map Reduce Paradigm

Input Files (wordCount.txt, mapper.py, reducer.py) :

```
hadoop@fedora:~/hadoop/experiments$ hadoop jar ../share/hadoop/mapreduce/hadoop-mapreduce-examples-3.4.0.jar wordcount /words.txt /output
2024-08-30 00:45:53,318 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-08-30 00:45:53,797 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1724955478554_0005
2024-08-30 00:45:54,195 INFO input.FileInputFormat: Total input files to process : 1
2024-08-30 00:45:54,351 INFO mapreduce.JobSubmitter: number of splits:1
2024-08-30 00:45:55,023 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1724955478554_0005
2024-08-30 00:45:55,023 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-08-30 00:45:55,235 INFO conf.Configuration: resource-types.xml not found
2024-08-30 00:45:55,235 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-08-30 00:45:55,350 INFO impl.YarnClientImpl: Submitted application application_1724955478554_0005
2024-08-30 00:45:55,411 INFO mapreduce.Job: The url to track the job: http://fedora:8088/proxy/application_1724955478554_0005/
2024-08-30 00:45:55,413 INFO mapreduce.Job: Running job: job_1724955478554_0005
2024-08-30 00:46:02,612 INFO mapreduce.Job: Job job_1724955478554_0005 running in uber mode : false
2024-08-30 00:46:02,615 INFO mapreduce.Job:  map 0% reduce 0%
2024-08-30 00:46:06,892 INFO mapreduce.Job:  map 100% reduce 0%
2024-08-30 00:46:12,942 INFO mapreduce.Job:  map 100% reduce 100%
2024-08-30 00:46:13,954 INFO mapreduce.Job: Job job_1724955478554_0005 completed successfully
2024-08-30 00:46:14,072 INFO mapreduce.Job: Counters: 54
```

$chmod 777 mapper.py reducer.py

$hadoop jar $HADOOP_STREAMING -input /exp2/word_count.txt -output
/exp2/output -mapper ~/DA-Lab/exp2/mapper.py -reducer ~/DA-
Lab/exp2/reducer.py

```
$ hadoop jar $HADOOP_STREAMING -input /exp1/word_count.txt -output /exp1/new_output -mapper ~/DA-Lab/mapper.py -reducer ~/DA-Lab/reducer.py
2024-08-28 03:19:55,058 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
packageJobJar: [/tmp/hadoop-unjar564064678185769582o/] [] /tmp/streamjob6319961755761707411.jar tmpDir=null
2024-08-28 03:19:58,598 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-08-28 03:19:59,548 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-08-28 03:20:02,347 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1724827394647_0002
2024-08-28 03:20:04,382 INFO mapred.FileInputFormat: Total input files to process : 1
2024-08-28 03:20:05,170 INFO mapreduce.JobSubmitter: number of splits:2
2024-08-28 03:20:06,083 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1724827394647_0002
2024-08-28 03:20:06,084 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-08-28 03:20:07,390 INFO conf.Configuration: resource-types.xml not found
2024-08-28 03:20:07,392 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-08-28 03:20:09,571 INFO impl.YarnClientImpl: Submitted application application_1724827394647_0002
2024-08-28 03:20:10,079 INFO mapreduce.Job: The url to track the job: http://kali:8088/proxy/application_1724827394647_0002/
2024-08-28 03:20:10,115 INFO mapreduce.Job: Running job: job_1724827394647_0002
2024-08-28 03:20:57,787 INFO mapreduce.Job: Job job_1724827394647_0002 running in uber mode : false
2024-08-28 03:20:57,794 INFO mapreduce.Job:  map 0% reduce 0%
2024-08-28 03:21:27,334 INFO mapreduce.Job:  map 100% reduce 0%
2024-08-28 03:21:50,144 INFO mapreduce.Job:  map 100% reduce 100%
```

$hdfs dfs -cat /output/*

```
hadoop@fedora:~/hadoop/experiments$ hdfs dfs -cat /output/*
Integer.parseInt()      1
String.valueOf()        1
and      1
char     1
convert 1
converting      1
int      2
method. 1
string  2
then     1
to       2
using    2
```

Result :

Thus a basic Word Count program is run using map-reduce.