EXP NO: 2   RUN A BASIC WORD COUNT MAP REDUCE PROGRAM TO UNDERSTAND MAP REDUCE PARADIGM

$mkdir DA-Lab
$cd DA-Lab
$mkdir exp2
$cd exp2
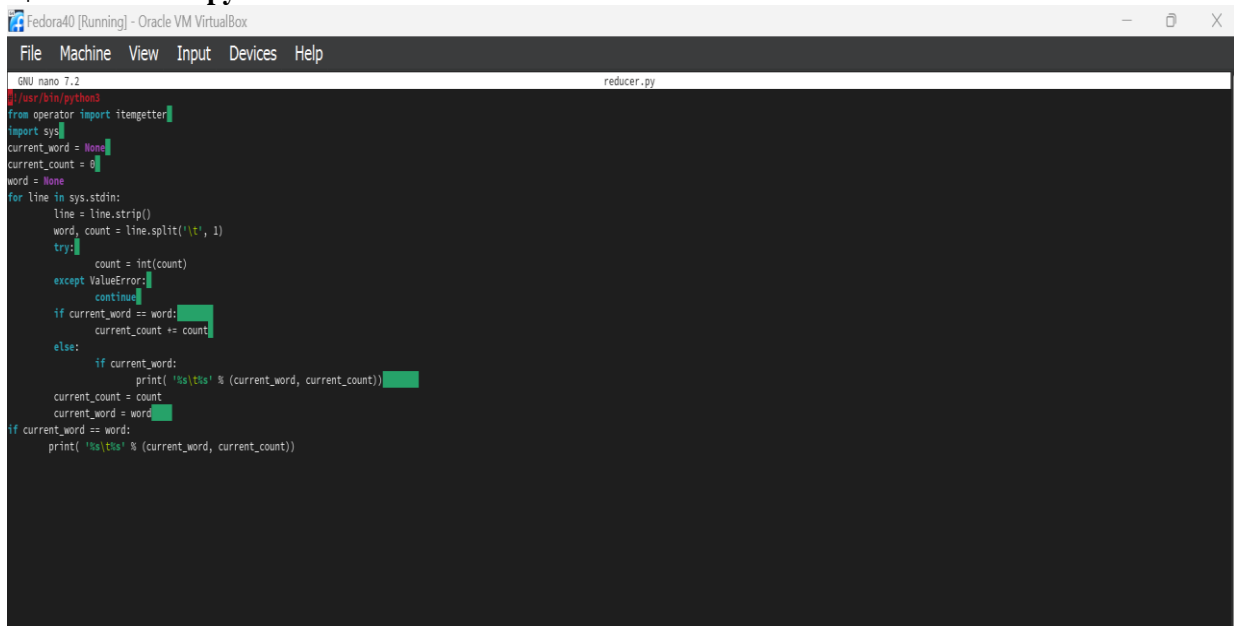
$nano word_count.txt

$nano mapper.py

$nano reducer.py

$start-all.sh

```
karthickragav@fedora:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as karthickragav in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [fedora]
Starting resourcemanager
Starting nodemanagers
```

$ jps

```
karthickragav@fedora:~$ jps
4916 Jps
4487 NodeManager
3609 NameNode
4009 SecondaryNameNode
4347 ResourceManager
3775 DataNode
```

$hdfs dfs -mkdir /exp2

$hdfs dfs -copyFromLocal ~/DA-Lab/exp2/word_count.txt /exp2

```
karthickragav@fedora:~/dalab/exp2$ hdfs dfs -ls /word_count_in_py
Found 2 items
-rw-r--r--   1 karthickragav supergroup         83 2024-09-01 21:13 /word_count_in_py/count.txt
drwxr-xr-x   - karthickragav supergroup          0 2024-09-01 21:14 /word_count_in_py/new_output
karthickragav@fedora:~/dalab/exp2$
```

$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

```
2024-09-01 21:14:26,332 INFO mapreduce.Job: Job job_local307751264_0001 completed successfully
2024-09-01 21:14:27,224 INFO mapreduce.Job: Counters: 36
        File System Counters
                FILE: Number of bytes read=283134
                FILE: Number of bytes written=1594029
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=166
                HDFS: Number of bytes written=92
                HDFS: Number of read operations=15
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=4
                HDFS: Number of bytes read erasure-coded=0
        Map-Reduce Framework
                Map input records=3
                Map output records=11
                Map output bytes=105
                Map output materialized bytes=133
                Input split bytes=100
                Combine input records=0
                Combine output records=0
                Reduce input groups=10
                Reduce shuffle bytes=133
                Reduce input records=11
                Reduce output records=10
                Spilled Records=22
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=206
                Total committed heap usage (bytes)=421527552
        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=83
        File Output Format Counters
                Bytes Written=92
2024-09-01 21:14:27,224 INFO streaming.StreamJob: Output directory: /word_count_in_py/new_output
karthickragav@fedora:~$ hdfs dfs -cat /word_count_in_py/new_output/part-00000
1          1
2          1
DA         1
experiment    1
experiments   1
hadoop 1
installation  1
lab    1
program 1
wordcount    1
karthickragav@fedora:~$
```

$hdfs dfs -cat /exp2/output/*

```
karthickragav@fedora:~/dalab/exp2$ hdfs dfs -cat /word_count_in_py/new_output/part-00000
1       1
2       1
DA      1
experiment      1
experiments     1
hadoop  1
installation    1
lab     1
program 1
wordcount       1
karthickragav@fedora:~/dalab/exp2$
```