

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

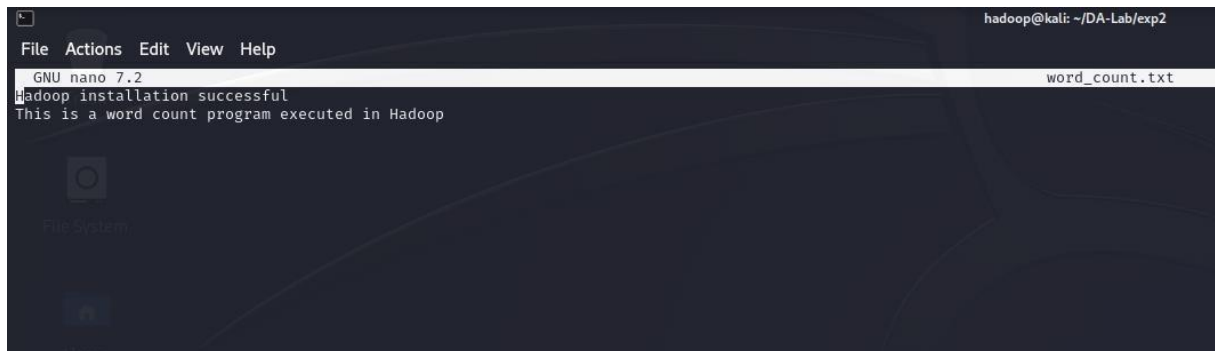
**\$mkdir DA-Lab**

**\$cd DA-Lab**

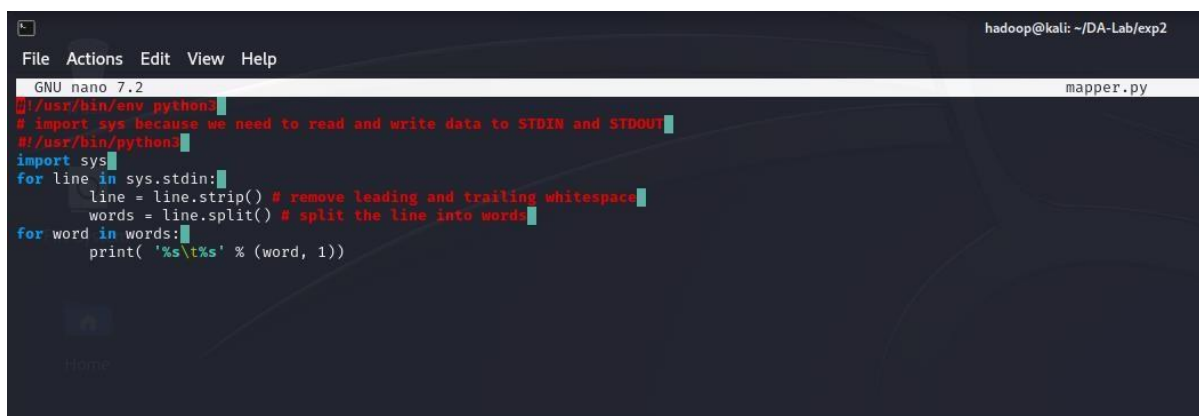
**\$mkdir exp2**

**\$cd exp2**

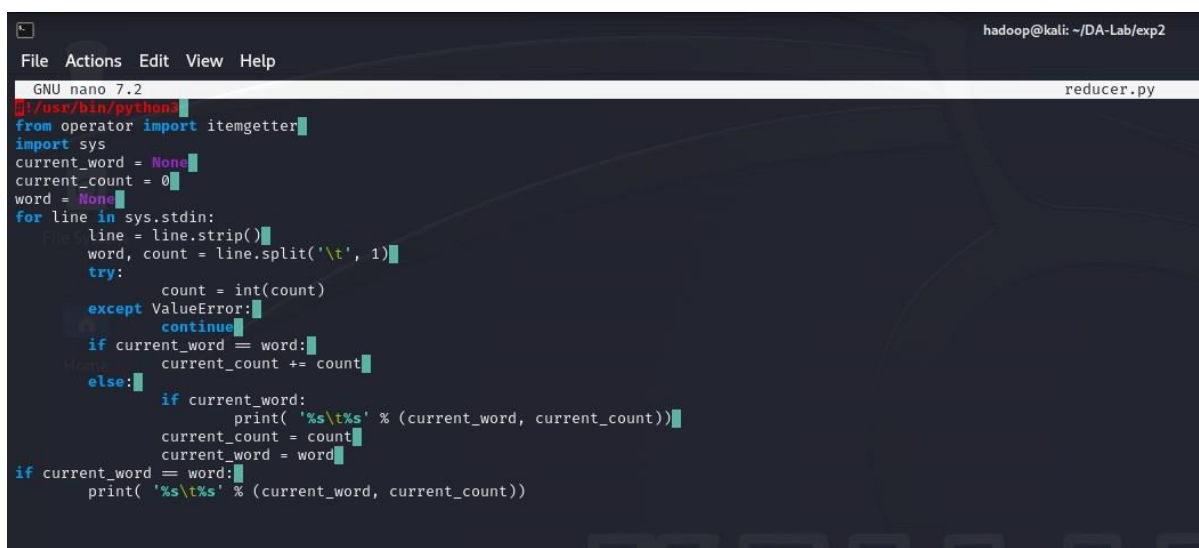
**\$nano word\_count.txt**

A screenshot of a terminal window showing the nano text editor. The editor is open to a file named 'word\_count.txt'. The text inside the file reads: 'Hadoop installation successful' followed by 'This is a word count program executed in Hadoop' on the next line. The terminal prompt is 'hadoop@kali: ~/DA-Lab/exp2'.

**\$nano mapper.py**

A screenshot of a terminal window showing the nano text editor. The editor is open to a file named 'mapper.py'. The code is a Python script for the mapper function. It imports sys, reads lines from stdin, strips whitespace, splits into words, and prints each word with a tab and a count of 1. The terminal prompt is 'hadoop@kali: ~/DA-Lab/exp2'.

**\$nano reducer.py**

A screenshot of a terminal window showing the nano text editor. The editor is open to a file named 'reducer.py'. The code is a Python script for the reducer function. It imports sys and operator, initializes current\_word and current\_count, and processes lines from stdin to calculate word counts. The terminal prompt is 'hadoop@kali: ~/DA-Lab/exp2'.

**\$start-all.sh**

```
hadoop@kali: ~  
File Actions Edit View Help  
  
hadoop@kali: ~  
$ start-all.sh  
WARNING: Attempting to start all Apache Hadoop daemons as hadoop 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 [kali]  
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true  
2024-09-11 04:59:16,429 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
Starting resourcemanager  
Starting nodemanagers
```

**\$ jps**

```
(hadoop@kali)-[~]
$ jps
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
14436 NodeManager
16772 Jps
13830 SecondaryNameNode
14311 ResourceManager
13597 DataNode
13471 NameNode

(hadoop@kali)-[~]
$
```

```
$hdfs dfs -mkdir /exp2
```

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

```
(hadoop@kali) [~/hadoop/bin]
$ ./hdfs dfs -ls /exp2
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings-on -Dswing.aatext=true
2024-09-21 00:05:07,404 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x   - hadoop supergroup          0 2024-09-13 01:05 /exp2/output
-rw-r--r--   1 hadoop supergroup        80 2024-09-13 01:02 /exp2/word_count.txt
```

### \$Schmod 777 mapper.py reducer.py

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

The image shows a Kali Linux virtual machine interface. At the top, the title bar reads "kali-linux-2023.4-vmware-vmware Workstation 17 Player (Non-commercial use only)". Below the title bar is a menu bar with "File", "Actions", "Edit", and "View" menus. The main window displays a terminal session with the following content:  

```
hadoop@kali:~/hadoop$  
File Actions Edit View Help  
getmerge SOURCE DEST  
put SOURCE DEST  
df [-h]  
truncate SIZE FILE  
  
[hadoop@kali:~]$  
$ 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  
2024-08-28 03:19:55,060 INFO client.DefaultHadoopFileProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032  
2024-08-28 03:19:55,060 INFO client.DefaultHadoopFileProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032  
2024-08-28 03:19:55,060 INFO mapreduce.JobResourceDisposer: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1724827394647_0002  
2024-08-28 03:19:55,060 INFO mapreduce.Job: Total input files to process: 1  
2024-08-28 03:19:55,060 INFO mapreduce.JobSubmitter: number of splits:2  
2024-08-28 03:19:55,060 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1724827394647_0002  
2024-08-28 03:19:55,060 INFO mapreduce.JobSubmitter: Executing with tokens: []  
2024-08-28 03:19:55,060 INFO conf.Configuration: resource-types.xml not found  
2024-08-28 03:19:55,060 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.  
2024-08-28 03:19:55,060 INFO mapreduce.Job: Submitting application application_1724827394647_0002  
2024-08-28 03:19:55,060 INFO mapreduce.Job: The url to track the job: http://kali:8088/proxy/application_1724827394647_0002/  
2024-08-28 03:19:55,060 INFO mapreduce.Job: Running job: job_1724827394647_0002  
2024-08-28 03:19:55,060 INFO mapreduce.Job: job_1724827394647_0002 running in user mode : false  
2024-08-28 03:19:55,060 INFO mapreduce.Job: map 0% reduce 0%  
2024-08-28 03:19:55,060 INFO mapreduce.Job: map 100% reduce 0%  
2024-08-28 03:19:55,060 INFO mapreduce.Job: map 100% reduce 100%  
2024-08-28 03:19:55,060 INFO mapreduce.Job: job_1724827394647_0002 completed successfully  
2024-08-28 03:19:55,060 INFO mapreduce.Job: Counters: 34  
File System Counters  
FILE: Number of bytes read=36  
FILE: Number of bytes written=936334  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=36  
HDFS: Number of bytes written=0  
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=2  
Data-local map tasks=2  
Total time spent by all maps in occupied slots (ms)=51128  
Total time spent by all reduces in occupied slots (ms)=19553  
Total time spent by all map tasks (ms)=51128  
Total time spent by all reduce tasks (ms)=19553  
Total vcore-milliseconds taken by all map tasks=51128  
Total vcore-milliseconds taken by all reduce tasks=19553  
Total megabyte-milliseconds taken by all map tasks=52350872  
Total megabyte-milliseconds taken by all reduce tasks=24822272  
Map-Reduce Framework
```

**\$hdfs dfs -cat /exp2/output/\***

```
(hadoop@kali)-[~/hadoop/bin]
$ ./hdfs dfs -cat /exp2/output/*
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-09-21 00:07:24,178 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
Hadoop 1
This 1
a 1
count 1
executed 1
in 1
is 1
program 1
word 1
```