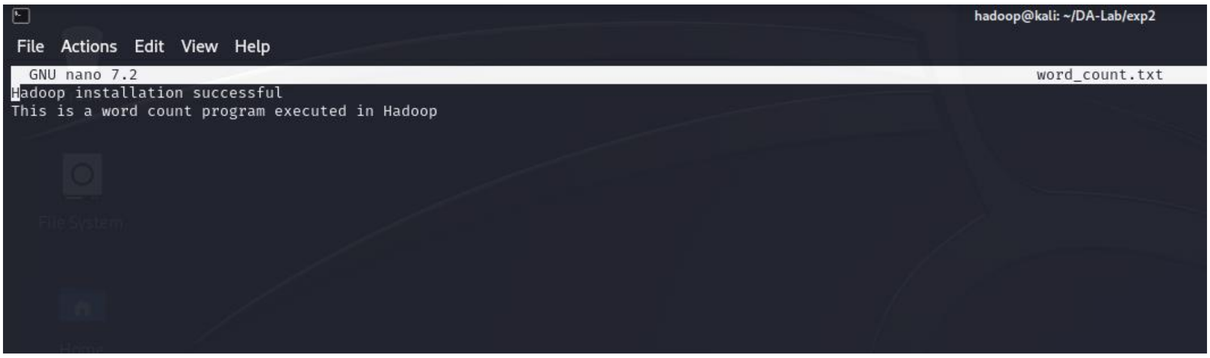
EXP NO: 2 RUN A BASIC WORD COUNT MAP REDUCE PROGRAM TO UNDERSTAND MAP REDUCE PARADIGM

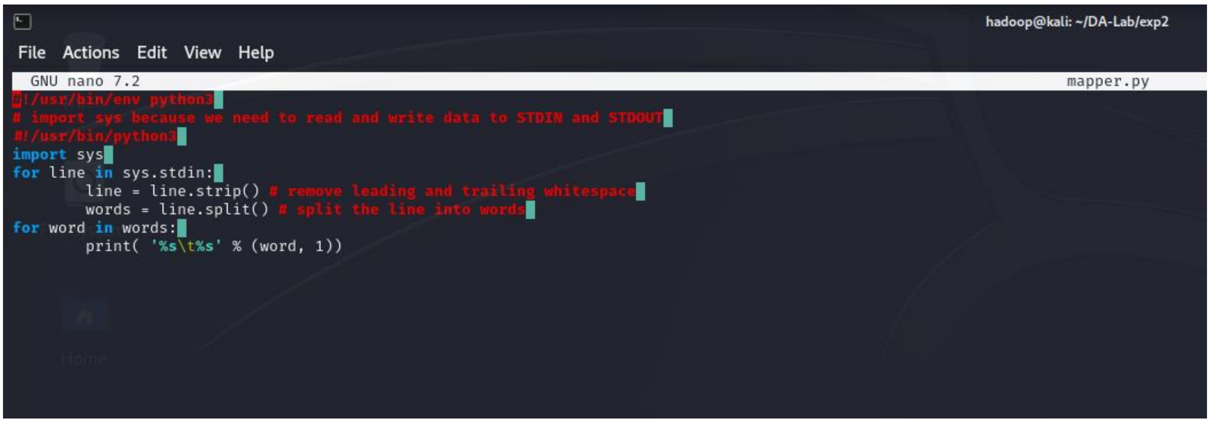
$mkdir DA-Lab
$cd DA-Lab
$mkdir exp2
$cd exp2

$nano word_count.txt



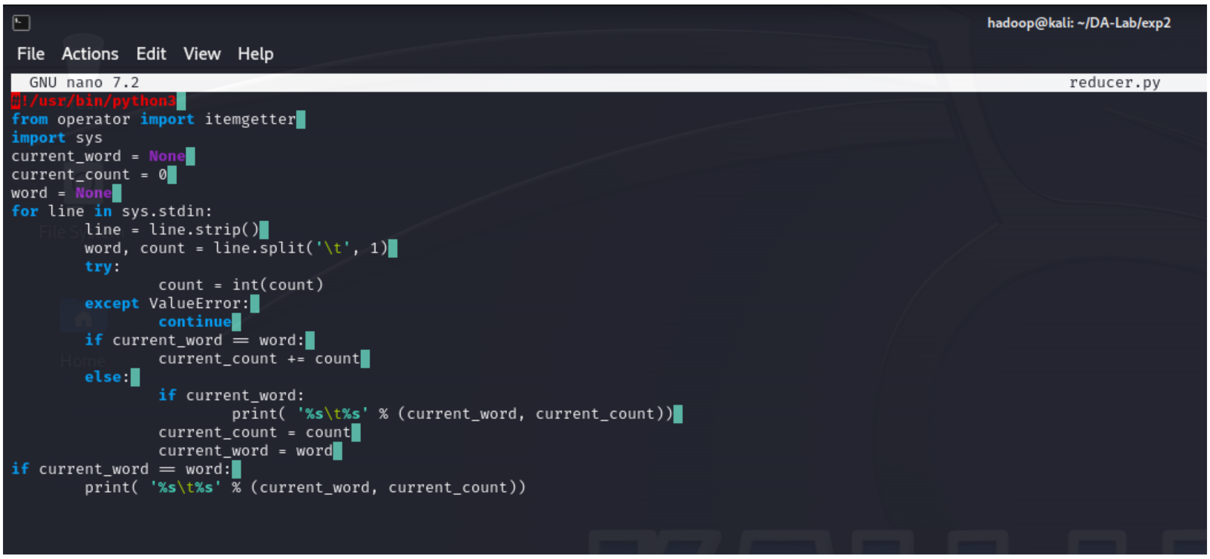
```
Hadoop installation successful
This is a word count program executed in Hadoop
```

$nano mapper.py



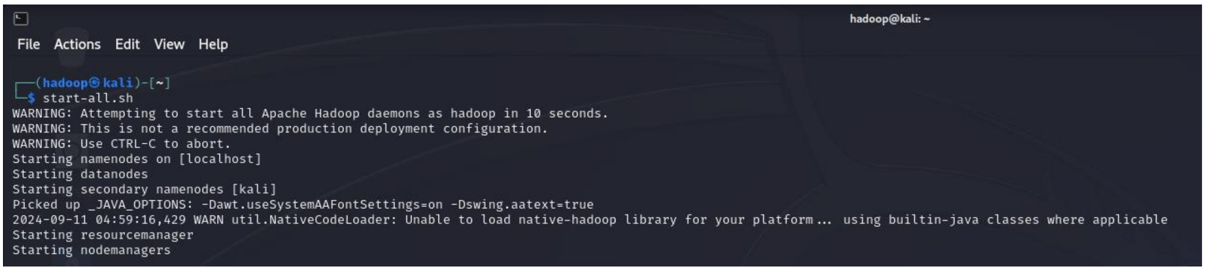
```
#!/usr/bin/env python3
# import sys because we need to read and write data to STDIN and STDOUT
#!/usr/bin/python3
import sys
for line in sys.stdin:
        line = line.strip() # remove leading and trailing whitespace
        words = line.split() # split the line into words
for word in words:
        print( '%s\t%s' % (word, 1))
```

$nano reducer.py



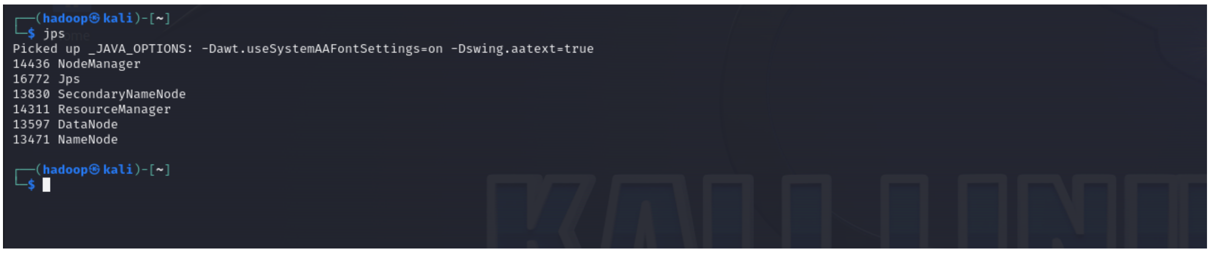
```
#!/usr/bin/python3
from operator import itemgetter
import sys
current_word = None
current_count = 0
word = None
for line in sys.stdin:
        line = line.strip()
        word, count = line.split('\t', 1)
        try:
                count = int(count)
        except ValueError:
                continue
        if current_word == word:
                current_count += count
        else:
                if current_word:
                        print( '%s\t%s' % (current_word, current_count))
                current_count = count
                current_word = word
if current_word == word:
        print( '%s\t%s' % (current_word, current_count))
```

$start-all.sh



```
┌──(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

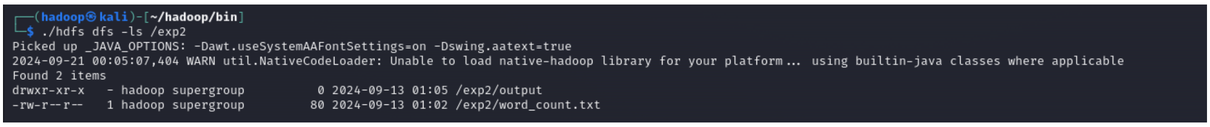


```
┌──(hadoop㉿kali)-[~]
└─$ jps
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
14436 NodeManager
16772 Jps
13830 SecondaryNameNode
14311 ResourceManager
13597 DataNode
13471 NameNode

┌──(hadoop㉿kali)-[~]
└─$ ▮
```

$hdfs dfs -mkdir /exp2

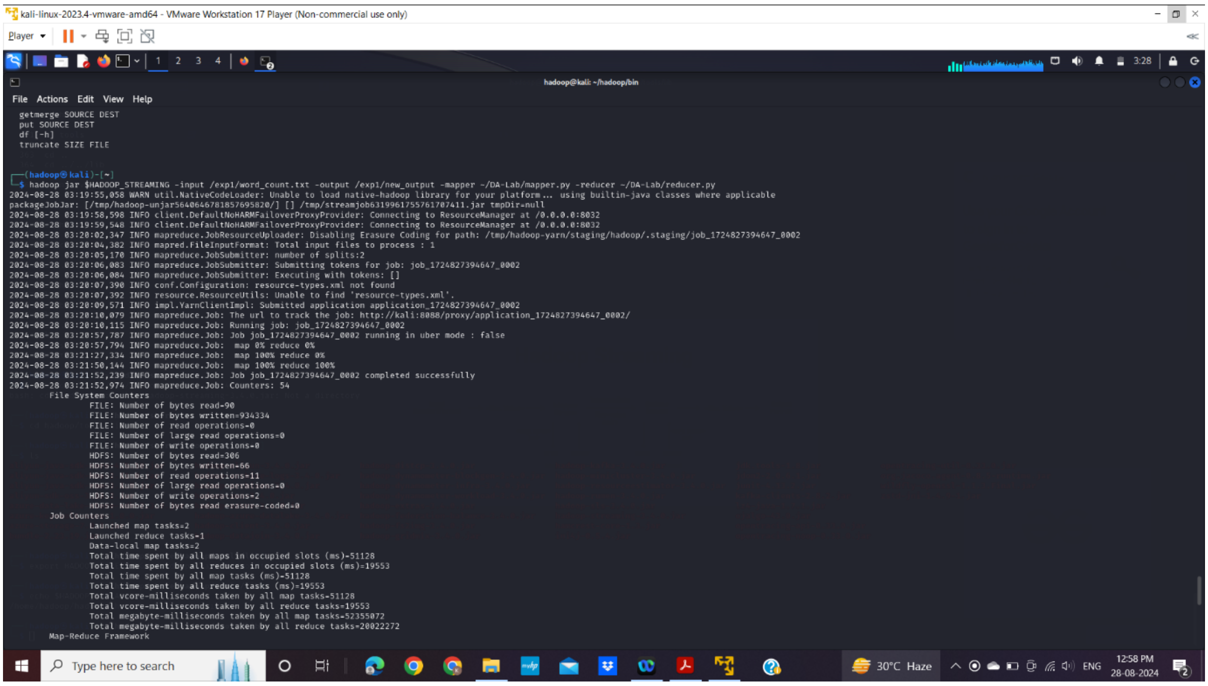
$hdfs dfs -copyFromLocal ~/DA-Lab/exp2/word_count.txt /exp2



```
┌──(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
```

$chmod 777 mapper.py reducer.py

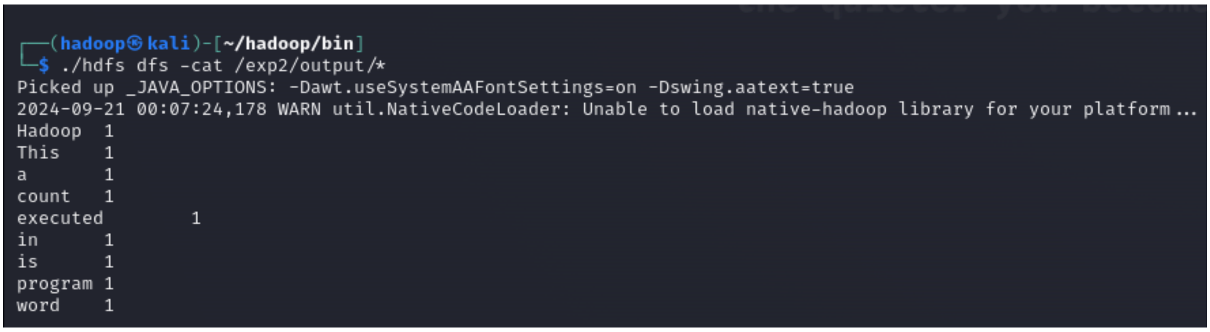
$hadoop jar $HADOOP_STREAMING -input /exp2/word_count.txt -output /exp2/output -mapper ~/DA-Lab/exp2/mapper.py -reducer ~/DA-Lab/exp2/reducer.py


kali-linux-2023.4-vmware-amd64 - VMware Workstation 17 Player (Non-commercial use only)
Player
1 2 3 4
3:28
hadoop@kali: ~/hadoop/bin
File Actions Edit View Help
Type here to search
30°C Haze
ENG
12:58 PM
28-08-2024

```
  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
packageJobJar: [/tmp/hadoop-unjar5640646781857695820/] [] /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%
2024-08-28 03:21:52,239 INFO mapreduce.Job: Job job_1724827394647_0002 completed successfully
2024-08-28 03:21:52,974 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=90
                FILE: Number of bytes written=934334
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=306
                HDFS: Number of bytes written=66
                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=1
                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=52355072
                Total megabyte-milliseconds taken by all reduce tasks=20022272
        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
```