

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

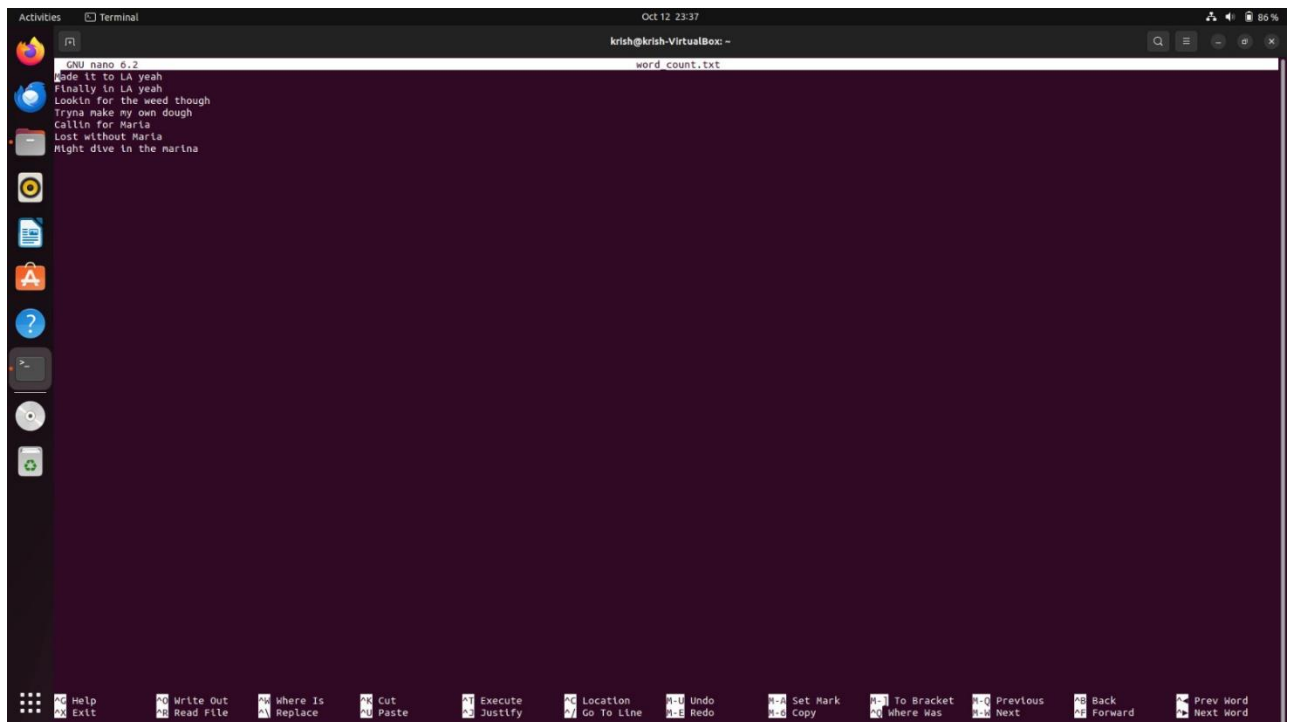
\$mkdir DA-Lab

\$cd DA-Lab

\$mkdir exp2

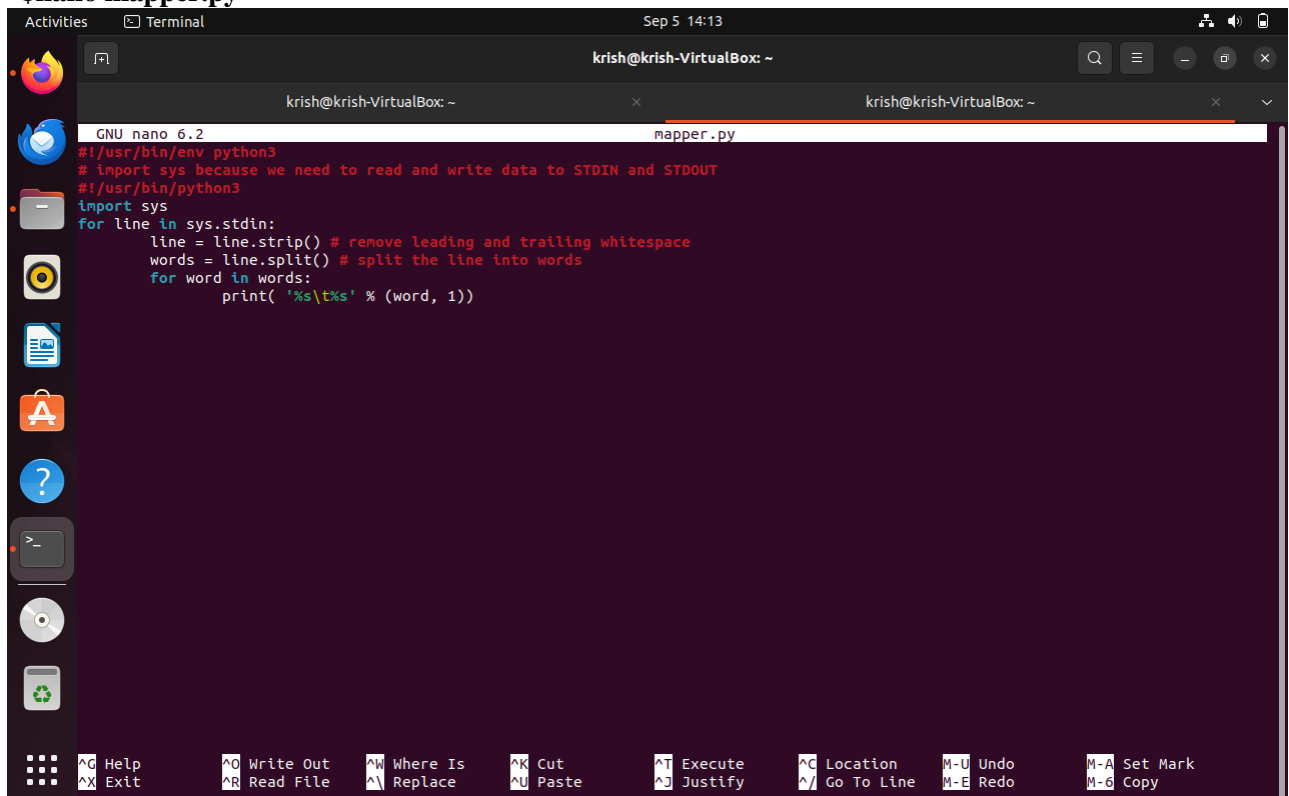
\$cd exp2

\$nano word_count.txt



```
GNU nano 6.2 word_count.txt
made it to LA yeah
Finally in LA yeah
Lookin for the weed though
Tryna make my own dough
Callin for Maria
Lost without Maria
Night drive in the marina
```

\$nano mapper.py



```
GNU nano 6.2 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

```

krish@krish-VirtualBox: ~
GNU nano 6.2 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))

```

Read 22 lines

Help Exit Write Out Read File Where Is Replace Cut Paste Execute Justify Location Go To Line Undo Redo Set Mark Copy

\$start-all.sh**\$ jps****\$hdfs dfs -mkdir /exp2****\$hdfs dfs -copyFromLocal ~/DA-Lab/exp2/word_count.txt /exp2**

```

krish@krish-VirtualBox: ~
krish@krish-VirtualBox:~$ nano word_count.txt
krish@krish-VirtualBox:~$ nano mapper.py
krish@krish-VirtualBox:~$ nano mapper.py
krish@krish-VirtualBox:~$
krish@krish-VirtualBox:~$ nano reducer.py
krish@krish-VirtualBox:~$ nano reducer.py
krish@krish-VirtualBox:~$ hdfsdfs -mkdir /word_count_in_python
hdfsdfs: command not found
krish@krish-VirtualBox:~$ hdfs dfs -mkdir /word_count_in_python
krish@krish-VirtualBox:~$ hdfsdfs -copyFromLocal /path/to/word_count.txt/word_count_in_python
hdfsdfs: command not found
krish@krish-VirtualBox:~$ hdfs dfs -copyFromLocal /path/to/word_count.txt/word_count_in_python
copyFromLocal: '.': No such file or directory: 'hdfs://localhost:9000/user/krish'
krish@krish-VirtualBox:~$ hdfs dfs -copyFromLocal /home/krish/word_count.txt/word_count_in_python
copyFromLocal: '.': No such file or directory: 'hdfs://localhost:9000/user/krish'
krish@krish-VirtualBox:~$ hdfs dfs -copyFromLocal ~/word_count.txt /word_count_in_python
krish@krish-VirtualBox:~$

```

\$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

```

krish@krish-VirtualBox: ~
krish@krish-VirtualBox: ~
krish@krish-VirtualBox: ~/hadoop-3.3.6/share...

krish@krish-VirtualBox:~$ hadoop jar ~/hadoop-3.3.6/share/hadoop/tools/lib/hadoop-streaming-3.3.6.jar -input /word_count_in_python/w
ord_count.txt -output /word_count_in_python/new_output -mapper ~/mapper.py -reducer ~/reducer.py
packageJobJar: [/tmp/hadoop-unjar5682300632999973691/] [] /tmp/streamjob8251200951483497453.jar tmpDir=null
2024-09-05 14:33:57,068 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-09-05 14:33:57,564 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-09-05 14:33:58,034 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/krish/.stagi
ng/job_1725524872355_0002
2024-09-05 14:33:58,621 INFO mapred.FileInputFormat: Total input files to process : 1
2024-09-05 14:33:59,532 INFO mapreduce.JobSubmitter: number of splits:2
2024-09-05 14:34:00,951 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1725524872355_0002
2024-09-05 14:34:00,951 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-09-05 14:34:01,471 INFO conf.Configuration: resource-types.xml not found
2024-09-05 14:34:01,472 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-09-05 14:34:02,416 INFO impl.YarnClientImpl: Submitted application application_1725524872355_0002
2024-09-05 14:34:02,613 INFO mapreduce.Job: The url to track the job: http://krish-VirtualBox:8088/proxy/application_1725524872355_0
002/
2024-09-05 14:34:02,619 INFO mapreduce.Job: Running job: job_1725524872355_0002
2024-09-05 14:34:28,789 INFO mapreduce.Job: Job job_1725524872355_0002 running in uber mode : false
2024-09-05 14:34:28,790 INFO mapreduce.Job: map 0% reduce 0%
2024-09-05 14:34:41,202 INFO mapreduce.Job: map 100% reduce 0%
2024-09-05 14:34:50,417 INFO mapreduce.Job: map 100% reduce 100%
2024-09-05 14:34:53,509 INFO mapreduce.Job: Job job_1725524872355_0002 completed successfully
2024-09-05 14:34:53,754 INFO mapreduce.Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=276
    FILE: Number of bytes written=836837
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=443
    HDFS: Number of bytes written=173
    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

```

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

```

krish@krish-VirtualBox: ~
krish@krish-VirtualBox: ~
krish@krish-VirtualBox: ~/hadoop-3.3.6/share...

krish@krish-VirtualBox:~$ hdfs dfs -cat /word_count_in_python/new_output/part-00000
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=225
File Output Format Counters
  Bytes Written=173
2024-09-05 14:34:53,754 INFO streaming.StreamJob: Output directory: /word_count_in_python/new_output
krish@krish-VirtualBox:~$ hdfs dfs -cat /word_count_in_python/new_output/part-00000
Callin 1
Finally 1
LA 2
Lookin 1
Lost 1
Made 1
Maria 2
Might 1
Tryna 1
dive 1
dough 1
for 2
in 2
it 1
make 1
marina 1
my 1
own 1
the 2
thought 1
to 1
weed 1
without 1
yeah 2
krish@krish-VirtualBox:~$

```