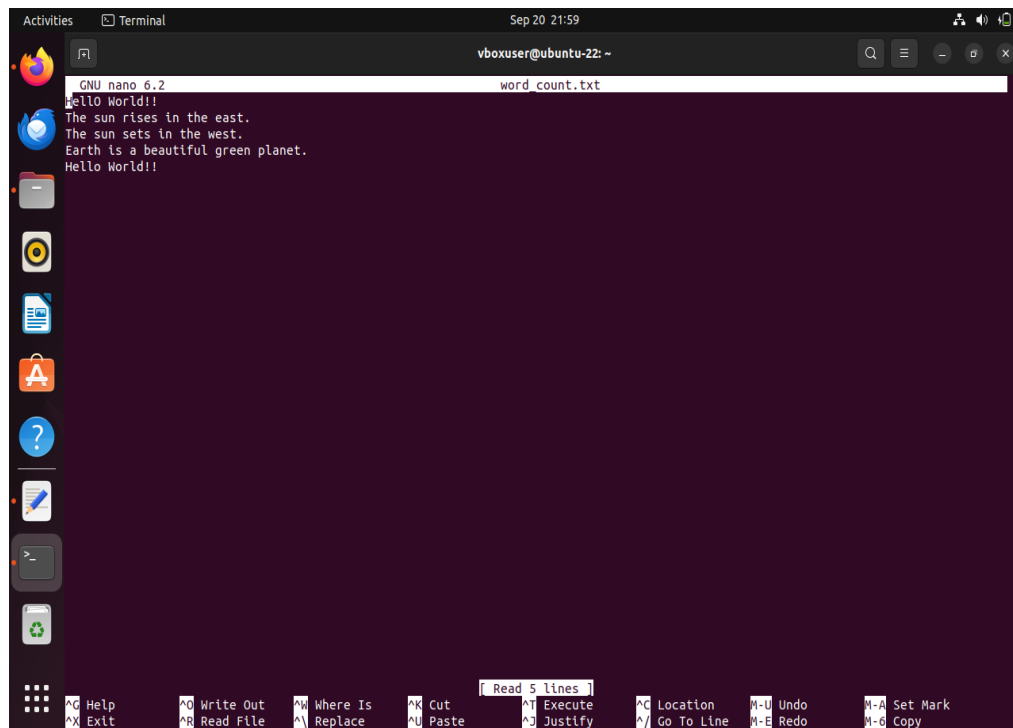**Exp.No: 2**

Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm

AIM:

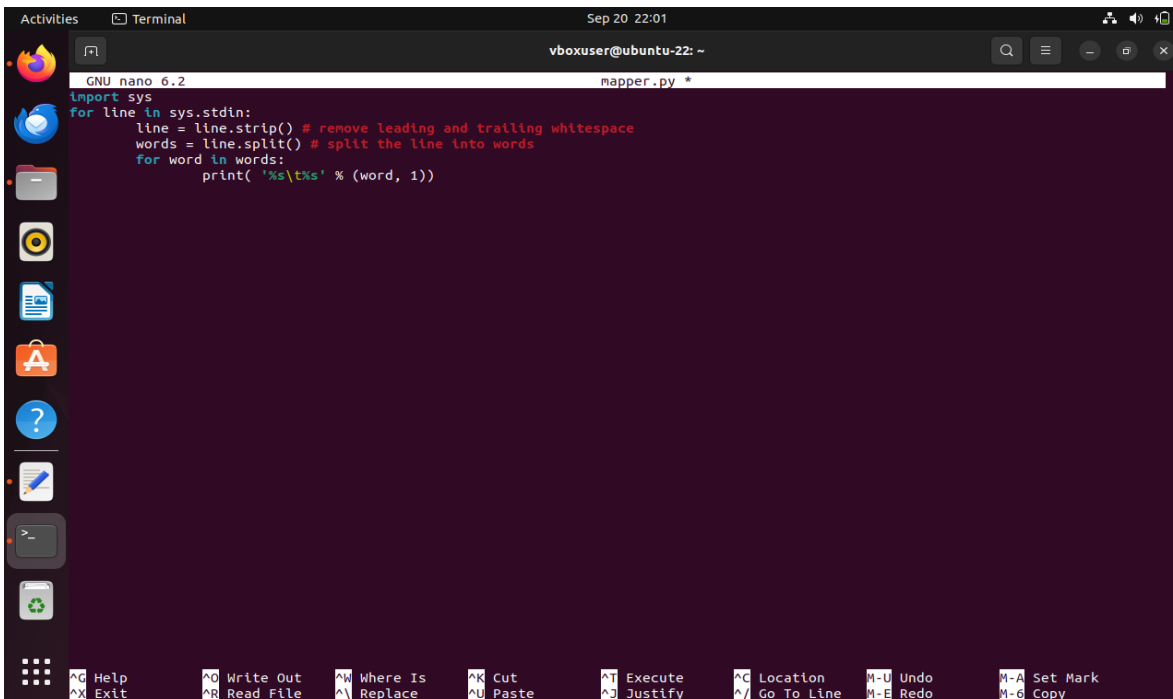To run a basic Word Count MapReduce program.

PROCEDURE:

Step 1: Create Data File: Create a file named "word_count_data.txt" and populate it with text data that you wish to analyse. Login with your hadoop user.

<p align="center">nano word_count.txt</p>

Step 2: Mapper Logic - mapper.py: Create a file named "mapper.py" to implement the logic for the mapper. The mapper will read input data from STDIN, split lines into words, and output each word with its count.
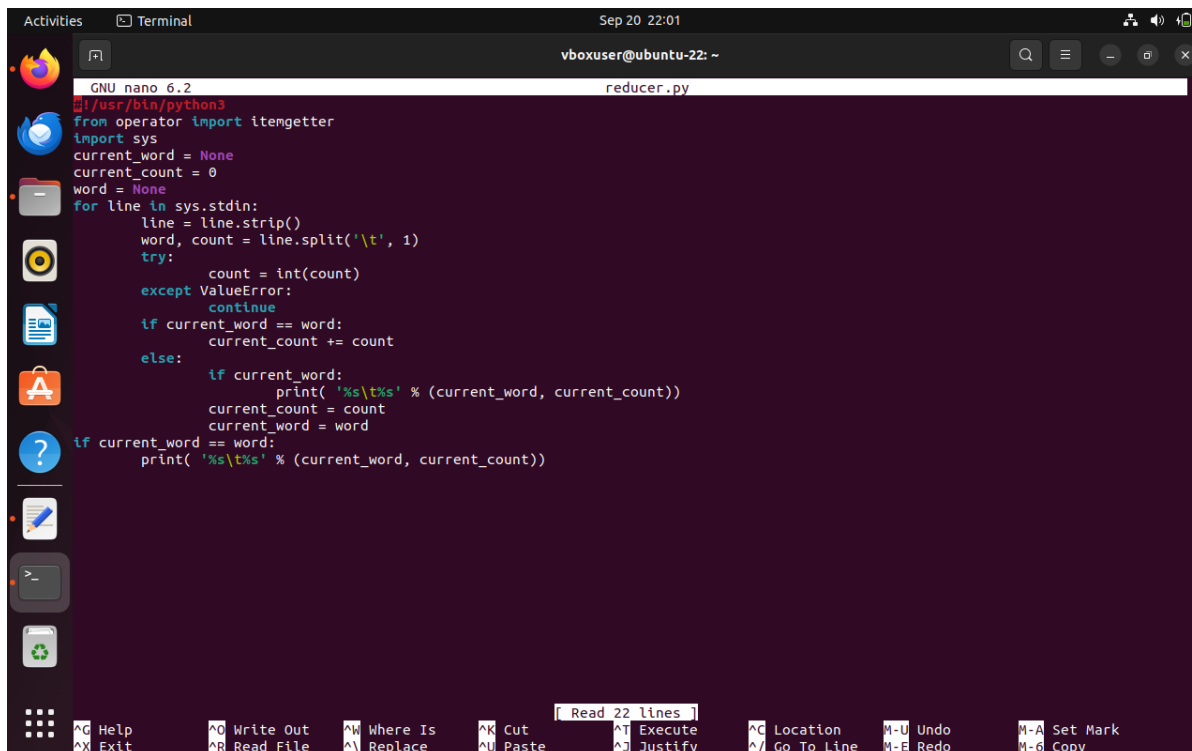
<p align="center">nano mapper.py</p>

Step 3: Reducer Logic - reducer.py: Create a file named "reducer.py" to implement the logic for the reducer. The reducer will aggregate the occurrences of each word and generate the final output.

nano reducer.py

Step 4: Prepare Hadoop Environment: Start the Hadoop daemons and create a directory in HDFS to store your data.

> **start-all.sh**
> **hdfs dfs -mkdir /word_count_in_python**
> **hdfs dfs -copyFromLocal /path/to/word_count.txt/word_count_in_python**

Step 5: Make Python Files Executable: Give executable permissions to your mapper.py and reducer.py files.

> **chmod 777 mapper.py reducer.py**

Step 6: Run Word Count using Hadoop Streaming: Download the latest hadoop-streaming jar file and place it in a location you can easily access. Then run the Word Count program using Hadoop Streaming.
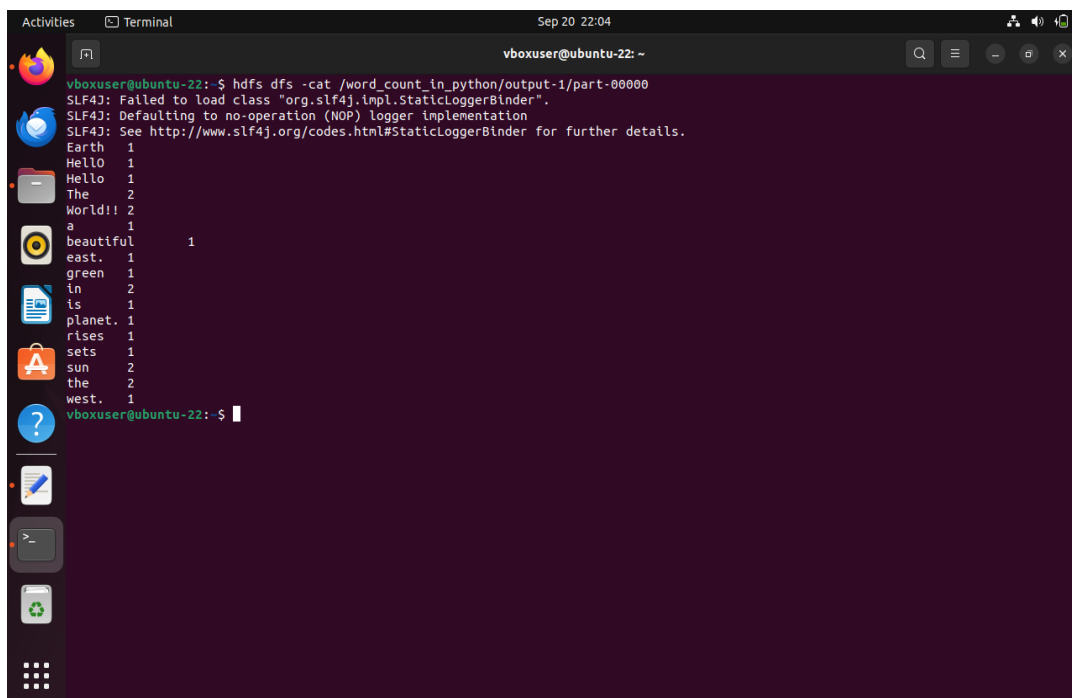
> **hadoop jar /path/to/hadoop-streaming-3.3.6.jar \ -input /word_count_in_python/word_count_data.txt \ -output /word_count_in_python/new_output \ -mapper /path/to/mapper.py \ -reducer /path/to/reducer.py**

Step 7: Check Output: Check the output of the Word Count program in the specified HDFS output directory.

> **hdfs dfs -cat /word_count_in_python/output-1/part-00000**

Step-8: Check the Output in the browser.

localhost:9870/explorer.html#/word_count_in_python/output-1

Block ID: 1073741871

Block Pool ID: BP-305058674-127.0.1.1-1726117174333

Generation Stamp: 1047

Size: 127

Availability:

- ubuntu-22.4.myguest.virtualbox.org

File contents

```
Earth     1
HellO     1
Hello     1
The  2
World!!   2
a    1
beautiful 1
east.     1
green     1
in   2
is   1
planet.   1
rises     1
sets 1
sun  2
the  2
west.     1
```

RESULT:

Thus, the program for basic Word Count Map Reduce has been executed successfully.