

Exp .No : 9

Roll no :210701258

DEMONSTRATE THE MAP REDUCE PROGRAMMING MODEL BY COUNTING THE NUMBER OF WORDS IN A FILE

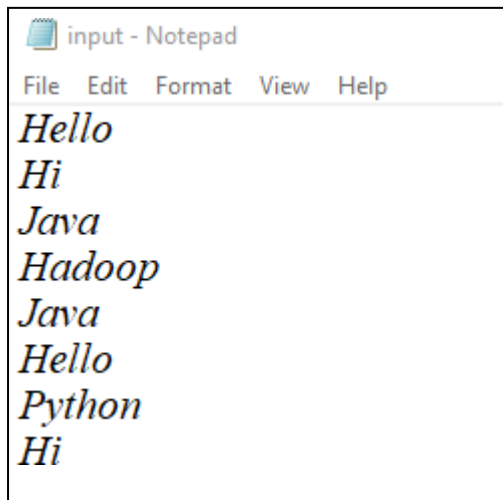
AIM:

To demonstrate the MAP REDUCE programming model for counting the number of words in a file.

PROCEDURE:

Step 1: Create Data File:

Create a file named "input.txt" and populate it with text data that you wish to analyse.



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.

mapper.py:

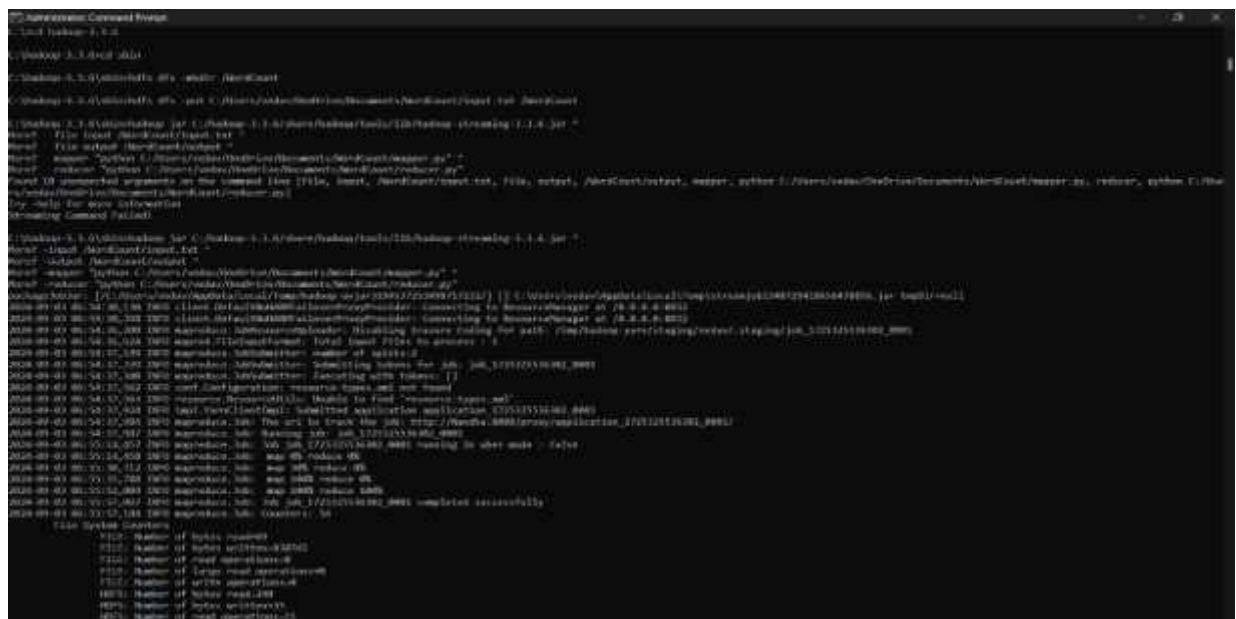
```
#!C:/Users/user/AppData/Local/Microsoft/WindowsApps/python.exe
import sys
for line in sys.stdin:
    line = line.strip()
    words = line.split()
    for word in words:
        print('%s\t%s'%(word,1))
```

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.

reducer.py:

```
#!C:/Users/user/AppData/Local/Microsoft/WindowsApps/python.exe
import sys
prev_word = None
prev_count = 0
for line in sys.stdin:
```

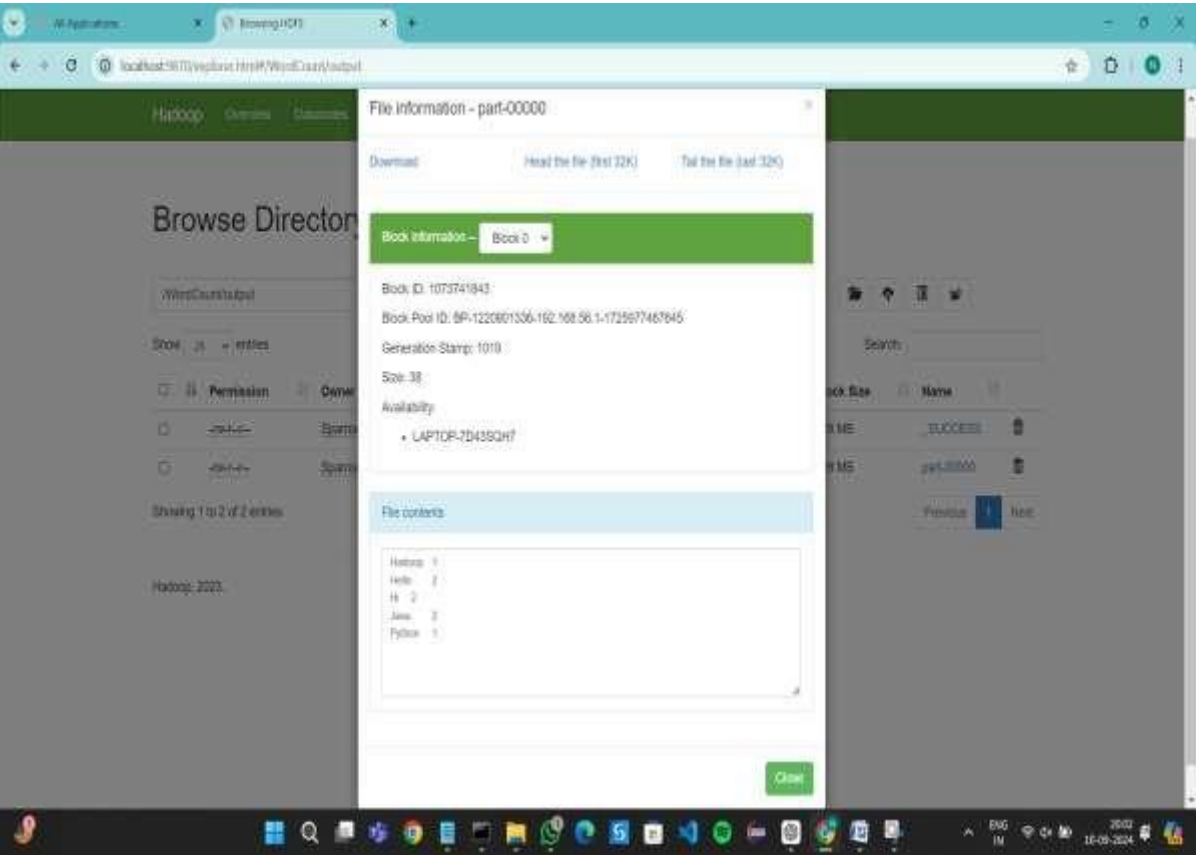


```
Administrator Command Prompt

Job: WordCount
Launched map tasks: 2
Launched reduce tasks: 1
Data-local map tasks: 2
Data-local map tasks: 2
Total time spent by all maps in occupied slots (ms): 22832
Total time spent by all reducers in occupied slots (ms): 54296
Total time spent by all map tasks (ms): 22832
Total time spent by all reduce tasks (ms): 54296
Total writeBytes (bytes) taken by all map tasks: 12922
Total writeBytes (bytes) taken by all reduce tasks: 12922
Total mapBytesWritten (bytes) taken by all map tasks: 12922
Total mapBytesWritten (bytes) taken by all reduce tasks: 12922

Map-Reduce Framework
Map input records: 2
Map output records: 2
Map output bytes: 12922
Map output materialized bytes: 12922
Input split bytes: 12922
Combine input records: 2
Combine output records: 2
Reduce input records: 2
Reduce shuffle bytes: 12922
Reduce input records: 2
Reduce output records: 2
Spilled Records: 2
Spilled Bytes: 12922
Hadoop MapOutputCommitter
GC Time elapsed (ms): 12922
CPU time spent (ms): 12922
Physical memory (bytes) allocated: 12922
Virtual memory (bytes) allocated: 12922
Total committed heap usage (bytes): 12922
Peak Map Physical memory (bytes): 12922
Peak Map Virtual memory (bytes): 12922
Peak Reduce Physical memory (bytes): 12922
Peak Reduce Virtual memory (bytes): 12922

Shuffle Errors
BAD_ID: 0
CONNECTION: 0
IO_ERROR: 0
MAPREDUCE: 0
NETIO: 0
OTHER: 0
FILE_IO_EXCEPTION: 0
File Output Format Counters
Bytes Written: 12922
2024-09-09 10:11:12,000 INFO org.apache.hadoop.mapreduce.lib.output.FileOutputFormat: Output directory: /home/centos/output
C:\Windows\system32\cmd.exe
```



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