

Exp .No : 9

Date : 30/08/2024

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

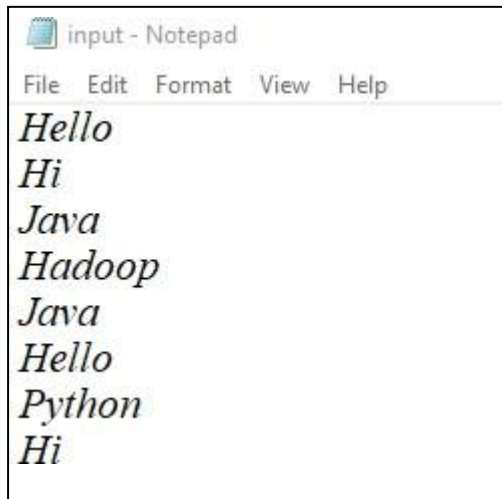
AIM:

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:
    line = line.strip()
    word, count = line.split('\t')
    count = int(count)
    if prev_word == word:
        prev_count += count
    else:
```



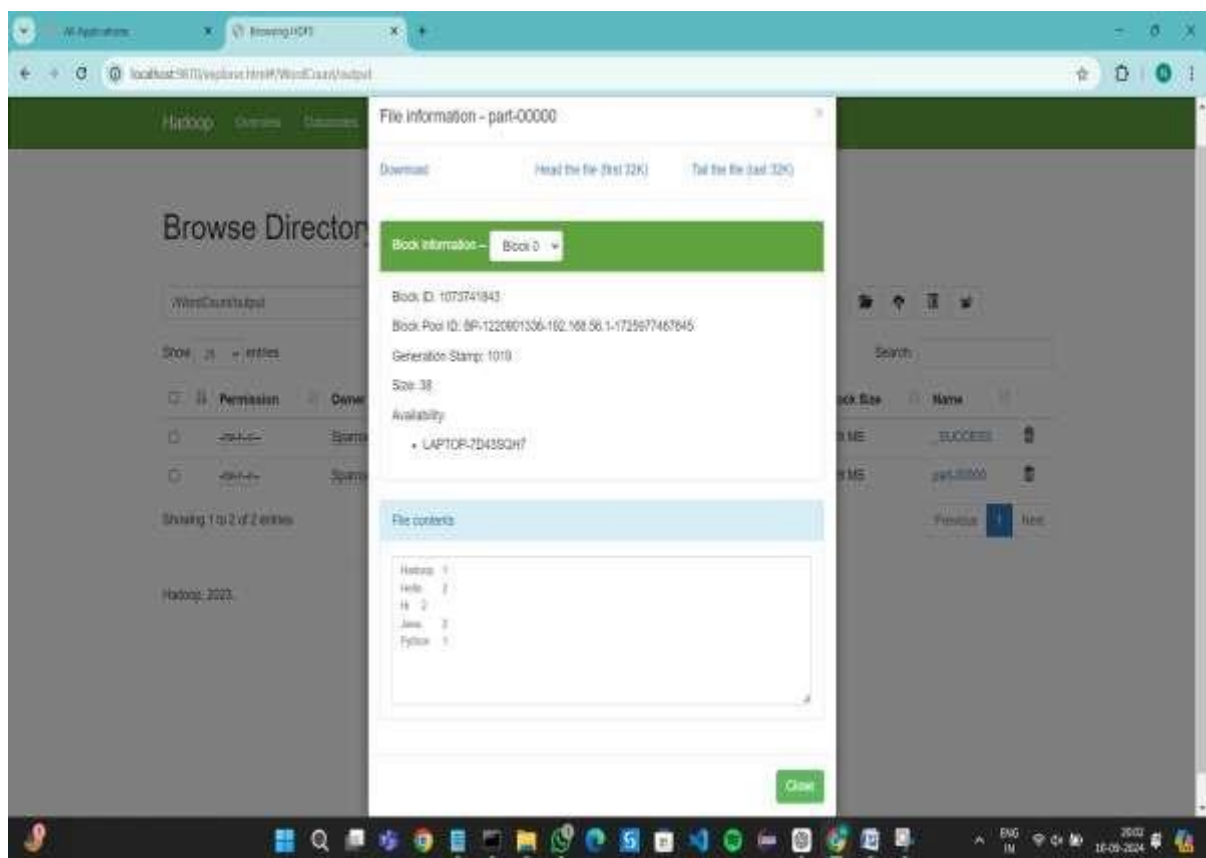
```
Administrator Command Prompt

bin\hadoop.cmd
Launched map task=0
Launched reduce task=0
Data read was finished
Total time spent by all maps in recorded slots (ms)=37332
Total time spent by all reducers in recorded slots (ms)=4539
Total time spent by all map tasks (ms)=37332
Total time spent by all reduce tasks (ms)=4539
Total write-milliseconds taken by all map tasks=27922
Total write-milliseconds taken by all reduce tasks=1404
Total mapwrite-milliseconds taken by all map tasks=2666738
Total mapwrite-milliseconds taken by all reduce tasks=1463639

Map Reduce Framework
Map input records=3
Map output records=3
Map output bytes=48
Map output materialized bytes=75
Input split bytes=128
Combine input records=4
Combine output records=4
Reduce input groups=0
Reduce shuffle bytes=0
Reduce task records=0
Reduce output records=0
Shuffled Map=0
Shuffled Map=0
Failed Shuffle=0
Failed Map shuffle=0
GC Time elapsed (ms)=587
CPU time spent (ms)=1801
Physical memory (bytes) committed=14745600
Virtual memory (bytes) committed=14745600
Total committed heap usage (bytes)=20777600
Peak Map Physical memory (bytes)=14745600
Peak Map Virtual memory (bytes)=14745600
Peak Reduce Physical memory (bytes)=20777600
Peak Reduce Virtual memory (bytes)=14745600

Shuffle Count
GEO=10-4
COMMIT=10-4
IO=10-4
WRITE=10-4
WRITE=10-4
WRITE=10-4
WRITE=10-4

File Input Format Counters
Byts Read=0
File Output Format Counters
Byts Written=0
2024-09-07 10:55:37,048 INFO chucwong.HadoopJob: Reduce directory: /user/hadoop/output
C:\Hadoop-3.3.6\bin>
```



RESULT:

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