

Exp .No : 9

Date :

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:
```

```
if prev_word == word:
```

Start the Hadoop daemons and create a directory in HDFS to store your data. Run the following commands to store the data in the WordCount Directory.

```
start-all.cmd cd C:/Hadoop/sbin hdfs dfs -mkdir /WordCount hdfs dfs -put
C:/Users/user/Documents/DataAnalytics/input.txt /WordCount hadoop jar
C:\hadoop\share\hadoop\tools\lib\hadoop-streaming-3.3.6.jar ^
```

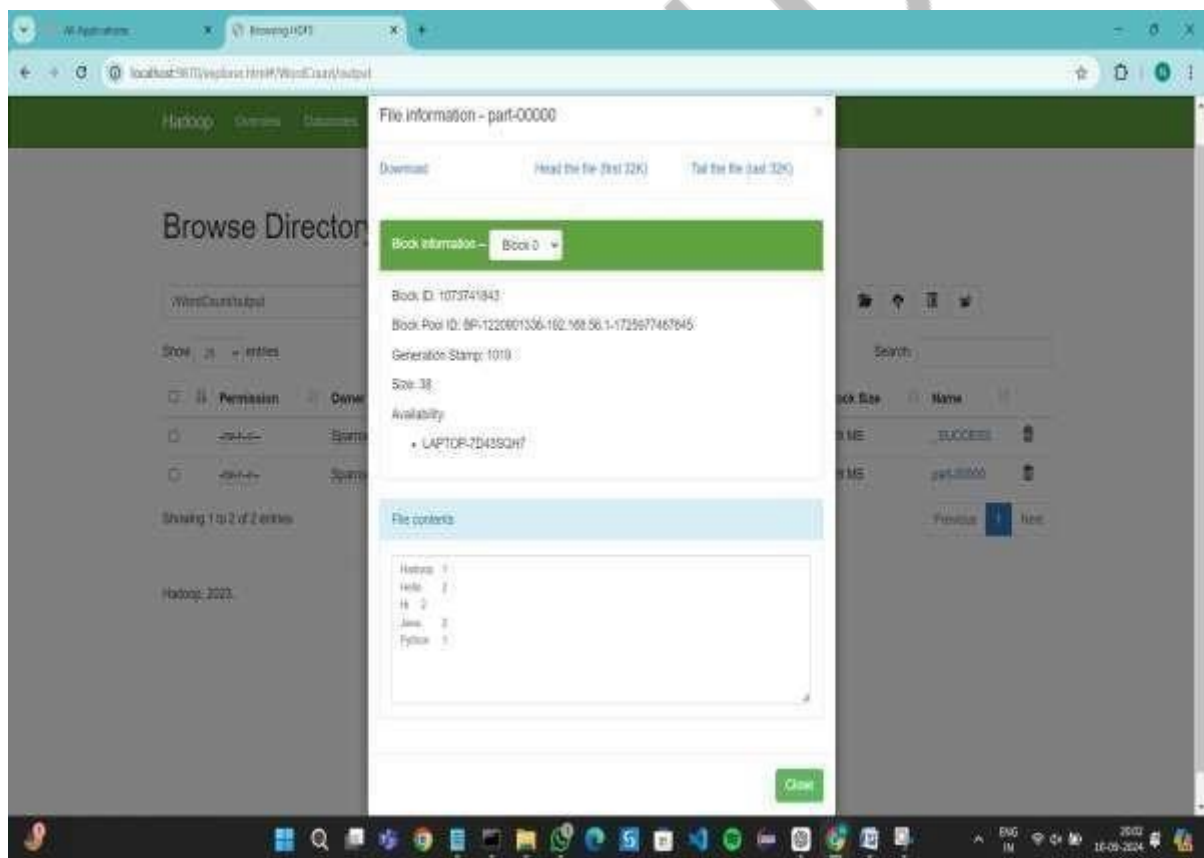
Check the output of the Word Count program in the specified HDFS output directory.

```
hdfs dfs -cat /WordCount/output/part-00000
```

OUTPUT:

[illegible]

```
AdminView: Command Prompt
Job Summary
Launched map tasks: 0
Launched reduce tasks: 0
Data Local S/N: 0/0
Total time spent by all maps in occupied slots (ms): 0/0
Total time spent by all reduces in occupied slots (ms): 0/0
Total time spent by all map tasks (ms): 0/0
Total time spent by all reduce tasks (ms): 0/0
Total shuffle bytes by all map tasks: 0/0
Total shuffle bytes by all reduce tasks: 0/0
Total map bytes written by all map tasks: 0/0
Total reduce bytes written by all reduce tasks: 0/0
Map-Reduce Summary
Map input records: 0
Map output records: 0
Map output bytes: 0
Map output materialized bytes: 0
Input split bytes: 0
Combine input records: 0
Combine output records: 0
Reduce input groups: 0
Reduce shuffle bytes: 0
Reduce input records: 0
Reduce output records: 0
Shuffled Map v2: 0
Failed Shuffle: 0
Maped Map outputs: 0
GC time elapsed (ms): 0/0
On time spent (ms): 0/0
Physical memory (bytes) assigned: 0/0
Virtual memory (bytes) assigned: 0/0
Total committed heap usage (bytes): 0/0
Peak Map Physical memory (bytes): 0/0
Peak Map Virtual memory (bytes): 0/0
Peak Reduce Physical memory (bytes): 0/0
Peak Reduce Virtual memory (bytes): 0/0
Shuffle Errors
SQS ID: 0
Connection: 0
IO error: 0
WARN: 0
WARN: 0
WARN: 0
WARN: 0
WARN: 0
File Input Format Counters
Bytes Read: 0
File Output Format Counters
Bytes Written: 0
2024-08-03 00:03:07,081 INFO org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter: Output directory: /user/hadoop/output
C:\Windows\system32\cmd.exe
```



RESULT:

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