Ex No: 2 **Roll No: 210701173**

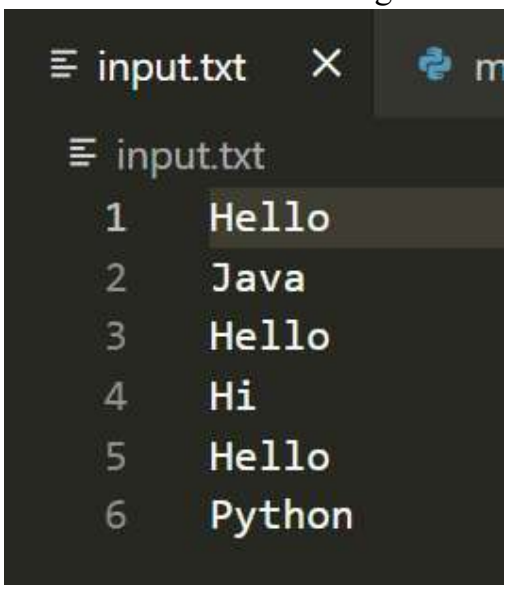
Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm.

AIM:

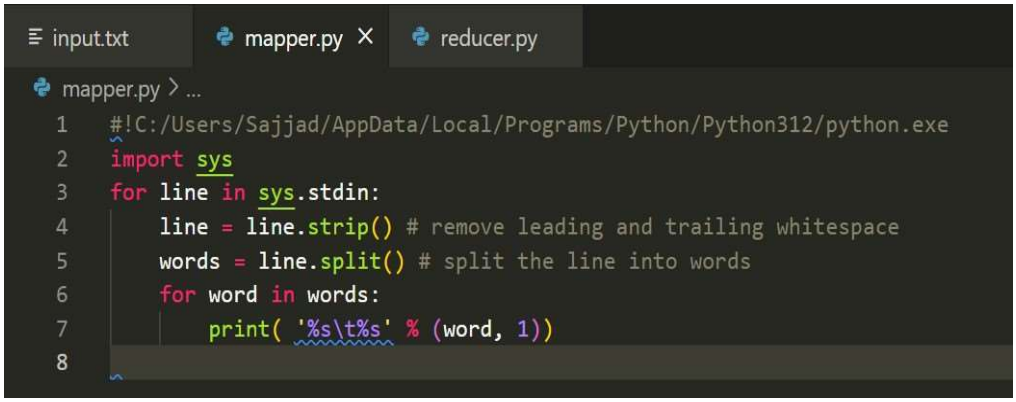
To run a basic Word Count MapReduce program using Hadoop.

PROCEDURE:

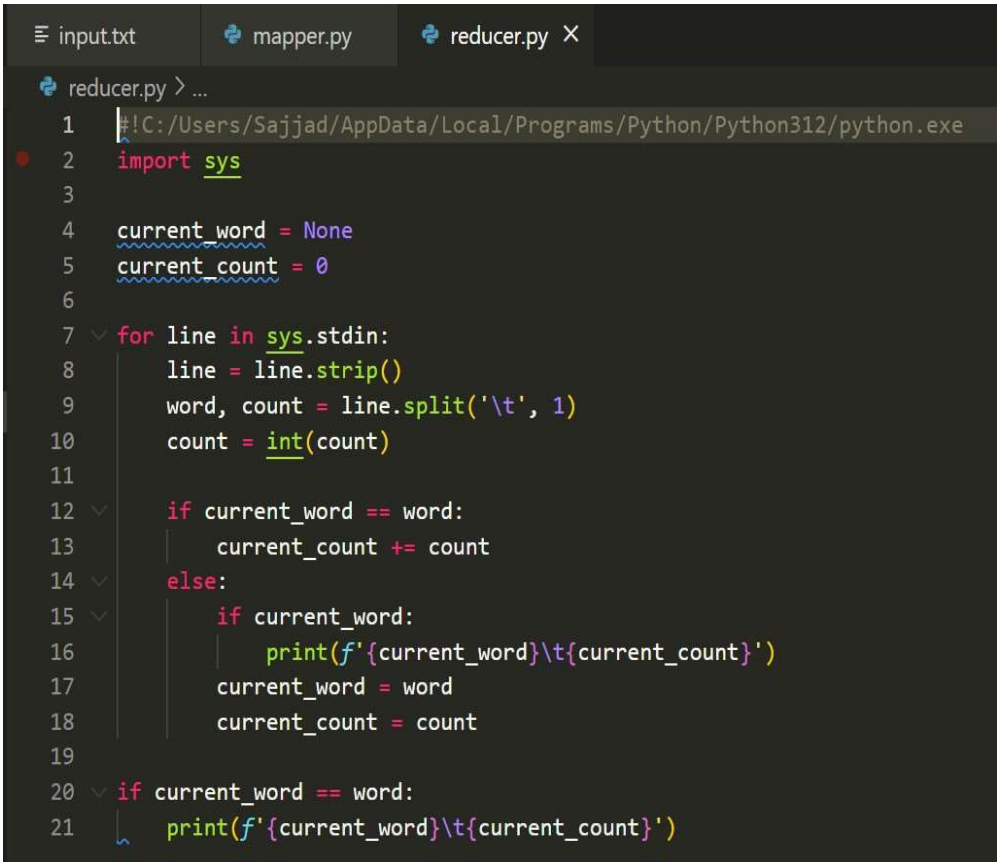
1. Create a text file containing necessary information.



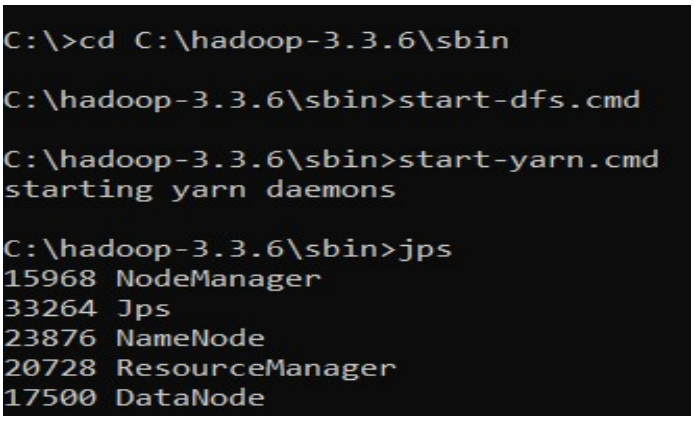
2. Create a mapper.py file that will read the input data and split lines into words.



3. Create a mapper.py file that will be used to implement the reducer.py file.



4. Start the Hadoop environment and create a directory to store values into HDFS.

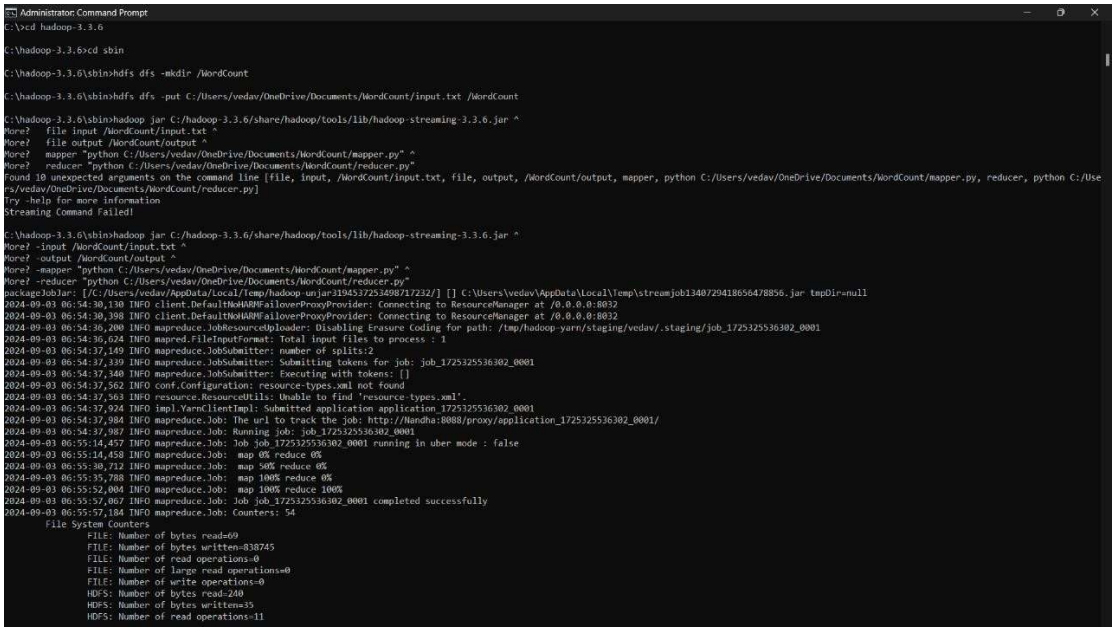


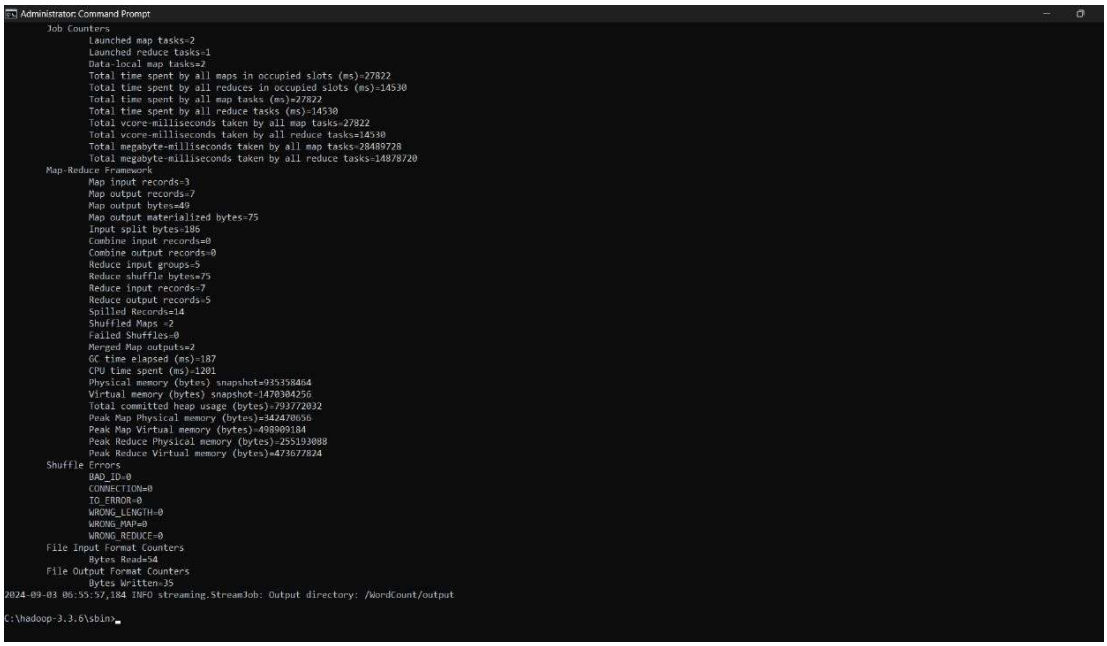
Commands to create directory, upload files and execute.

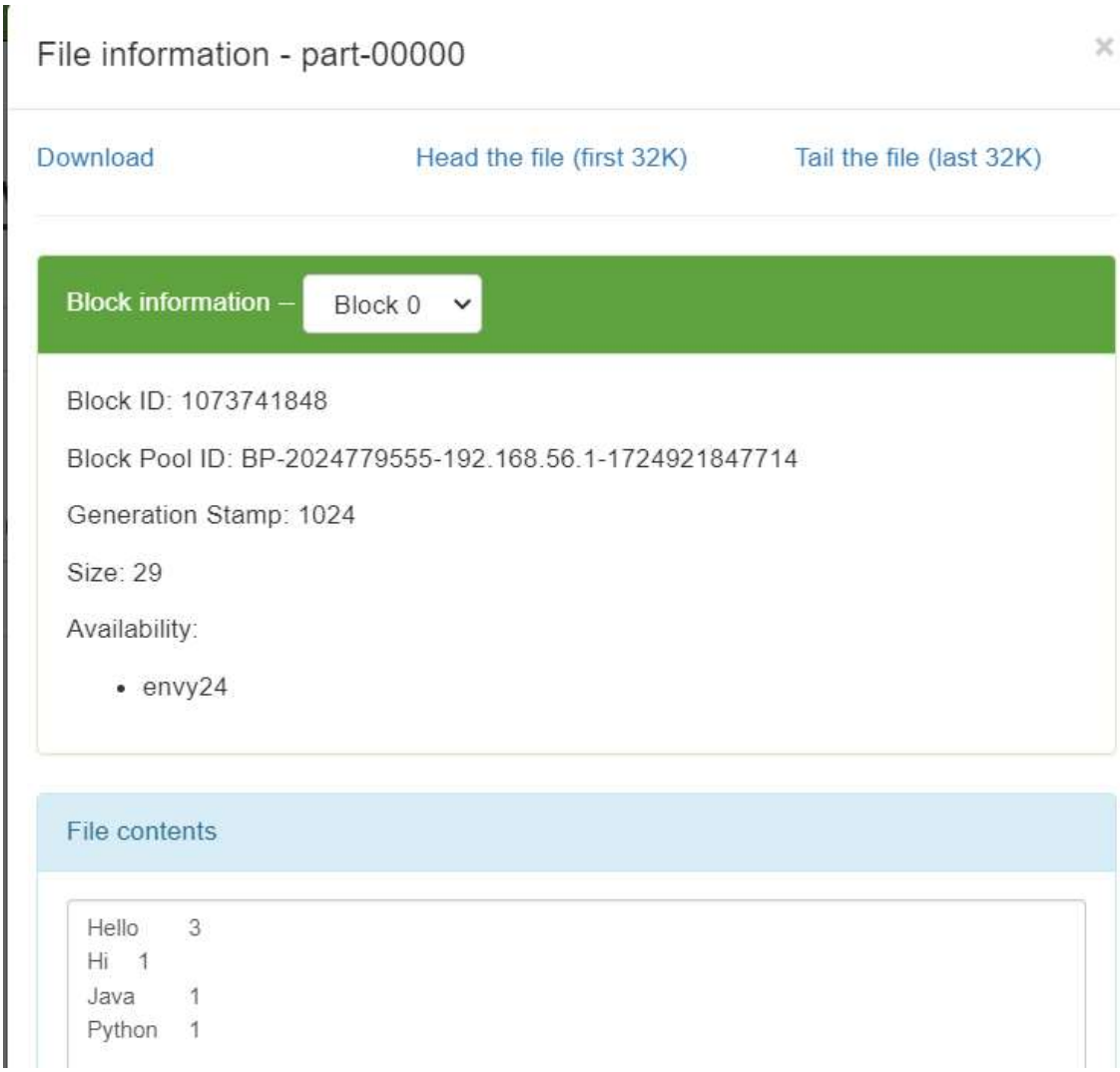
```
hdfs dfs -mkdir /WordCount

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 ^

-input /WordCount/input.txt ^

-output /WordCount/output ^

-mapper "python C:/ Users/user/Documents/DataAnalytics/mapper.py" ^

-reducer "python C:/ Users/user/Documents/DataAnalytics/reducer.py"
```

5. Check the output of the Word Count program in the specified HDFS output directory.
   hdfs dfs -cat /WordCount/output/part-00000

OUTPUT:




Administrator: Command Prompt
Job Counters
Launched map tasks=2
Launched reduce tasks=1
Data-local map tasks=2
Total time spent by all maps in occupied slots (ms)=27822
Total time spent by all reduces in occupied slots (ms)=14530
Total time spent by all map tasks (ms)=27822
Total time spent by all reduce tasks (ms)=14530
Total vcore-milliseconds taken by all map tasks=27822
Total vcore-milliseconds taken by all reduce tasks=14530
Total megabyte-milliseconds taken by all map tasks=28489728
Total megabyte-milliseconds taken by all reduce tasks=14878720
Map-Reduce Framework
Map input records=3
Map output records=7
Map output bytes=49
Map output materialized bytes=75
Input split bytes=186
Combine input records=0
Combine output records=0
Reduce input groups=5
Reduce shuffle bytes=75
Reduce input records=7
Reduce output records=5
Spilled Records=14
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=187
CPU time spent (ms)=1201
Physical memory (bytes) snapshot=935358464
Virtual memory (bytes) snapshot=1470304256
Total committed heap usage (bytes)=793772032
Peak Map Physical memory (bytes)=342470656
Peak Map Virtual memory (bytes)=498909184
Peak Reduce Physical memory (bytes)=255193088
Peak Reduce Virtual memory (bytes)=473677824
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=54
File Output Format Counters
Bytes Written=35
2024-09-03 06:55:37,184 INFO streaming.StreamJob: Output directory: /WordCount/output
C:\hadoop-3.3.6\sbin>



RESULT:

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