EX.NO. 2 210701159

IMPLEMENT WORD COUNT/FREQUENCY PROGRAMS USING MAPREDUCE

AIM:

To implement the python mapper and reducer programs using MapReduce to count the words in a text file using Hadoop.

PROCEDURE:

1. Open command prompt as administrator and start the Hadoop by using the command:

start-all.cmd

2. Create a new directory in the Hadoop file systems using the command:

hadoop fs -mkdir /wordCount

3. Upload the input text file into the wordCount directory using the command:

hadoop fs -put C:/DA/wordcount/input.txt /wordCount

- 4. Create the mapper and reducer files.
- 5. To execute the files with Hadoop streaming run the following command:

hadoop jar "C:\hadoop-3.3.6\share\hadoop\tools\lib\hadoop-streaming-3.3.6.jar" ^-input /wordCount/input.txt ^-output /wordCount/output/ ^-mapper "python C:/wordcount/mapper.py" ^-reducer "python C:/wordcount/reducer.py"

MAPPER.PY

```
\#!C:/Users/md\_aa/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))
```

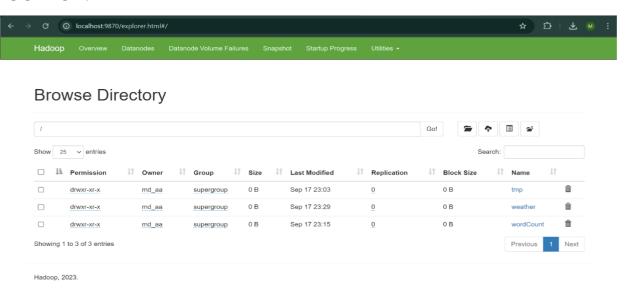
EX.NO. 2 210701159

REDUCER.PY

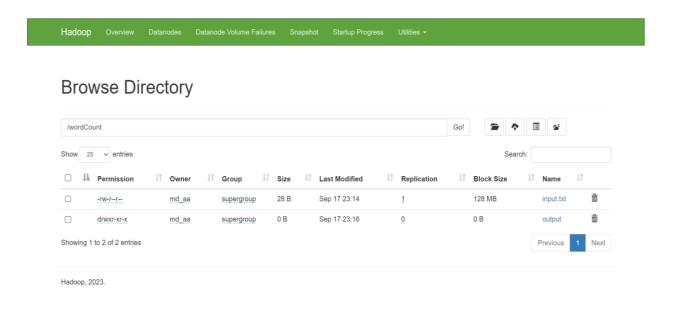
```
WindowsApps/python.exe
import sys
prev_word = None
prev\_count = 0
for line in sys.stdin:
  line = line.strip()
  word, count = line.split('\t')
  count = int(count)
  if(prev_word == word):
    prev_count += count
  else:
    if prev_word:
       print('%s\t%s' % (prev_word, prev_count))
    prev_count = count
    prev_word = word
if prev_word == word:
       print('%s\t%s' % (prev_word, prev_count))
```

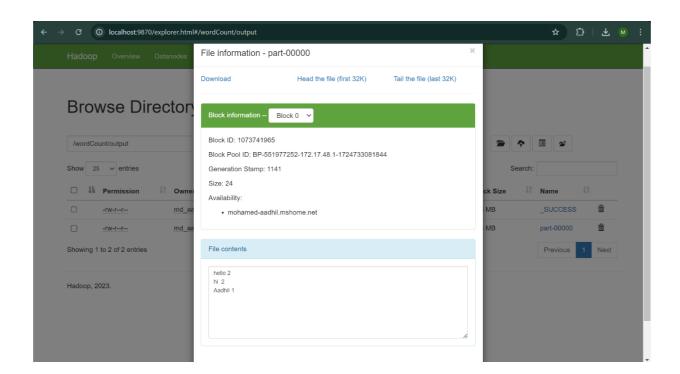
#!C:/Users/md_aa/AppData/Local/Microsoft/

OUTPUT:



EX.NO. 2 210701159





RESULT:

Thus, the implementation of the python mapper and reducer programs using MapReduce tocount the words in a text file using Hadoop is executed successfully.