EX.NO. 2 210701510

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:/Users/mercy/OneDrive/Documents/DataAnalytics/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 ^ -file C:/Users/mercy/Documents/DataAnalytics/mapper.py ^ -file C:/Users/mercy/Documents/DataAnalytics/reducer.py ^ -input /wordCount/input.txt ^ -output /user/output ^ -mapper "python mapper.py" ^ -reducer "python reducer.py"

MAPPER.PY

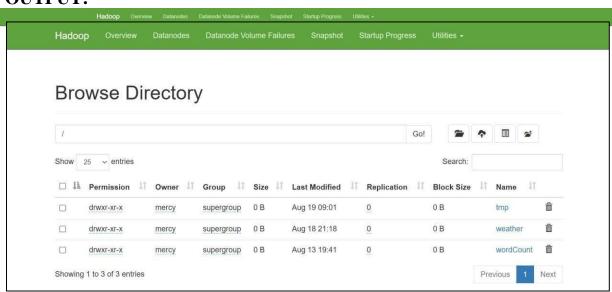
```
#!C:/ProgramData/chocolatey/bin/python3.exe
import sys
for line in sys.stdin:
    line = line.strip()
    words = line.split()
    for word in words:
        print('%s\t%s' % (word, 1))
```

EX.NO. 2

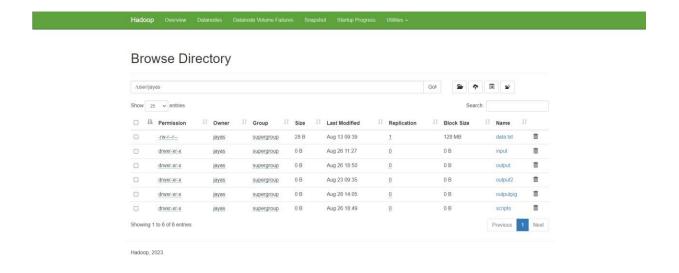
REDUCER.PY

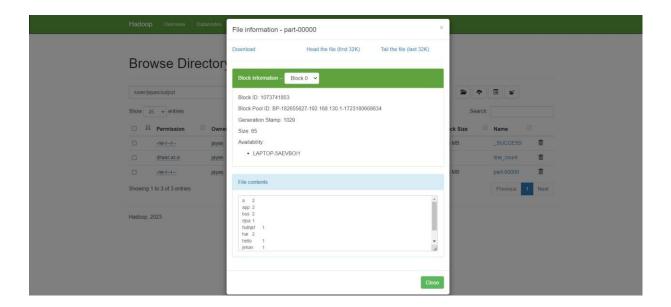
```
#!C:/ProgramData/chocolatey/bin/python3.exe
import sys
prev_word = None
prev\_count = 0
for line in sys.stdin:
  line = line.strip()
  word, count = line.split('\t')
  count = int(count)
  if(prev_word == word):
     prev_count += count
  else:
    if prev_word:
       print('%s\t%s' % (prev_word, prev_count))
     prev_count = count
    prev_word = word
if prev_word == word:
       print('%s\t%s' % (prev_word, prev_count))
```

OUTPUT:



EX.NO. 2





RESULT:

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