**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

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

hadoop fs -mkdir /wordCount

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

hadoop fs -put C:/Users/gjega/OneDrive/Documents/DataAnalytics/input.txt /wordcount

1. Create the mapper and reducer files.
2. 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/gjega/Documents/Hadoop\_wordcount/mapper.py ^ -file

C:/Users/gjega/Documents/Hadoop\_wordcount/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))

**REDUCER.PY**

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:

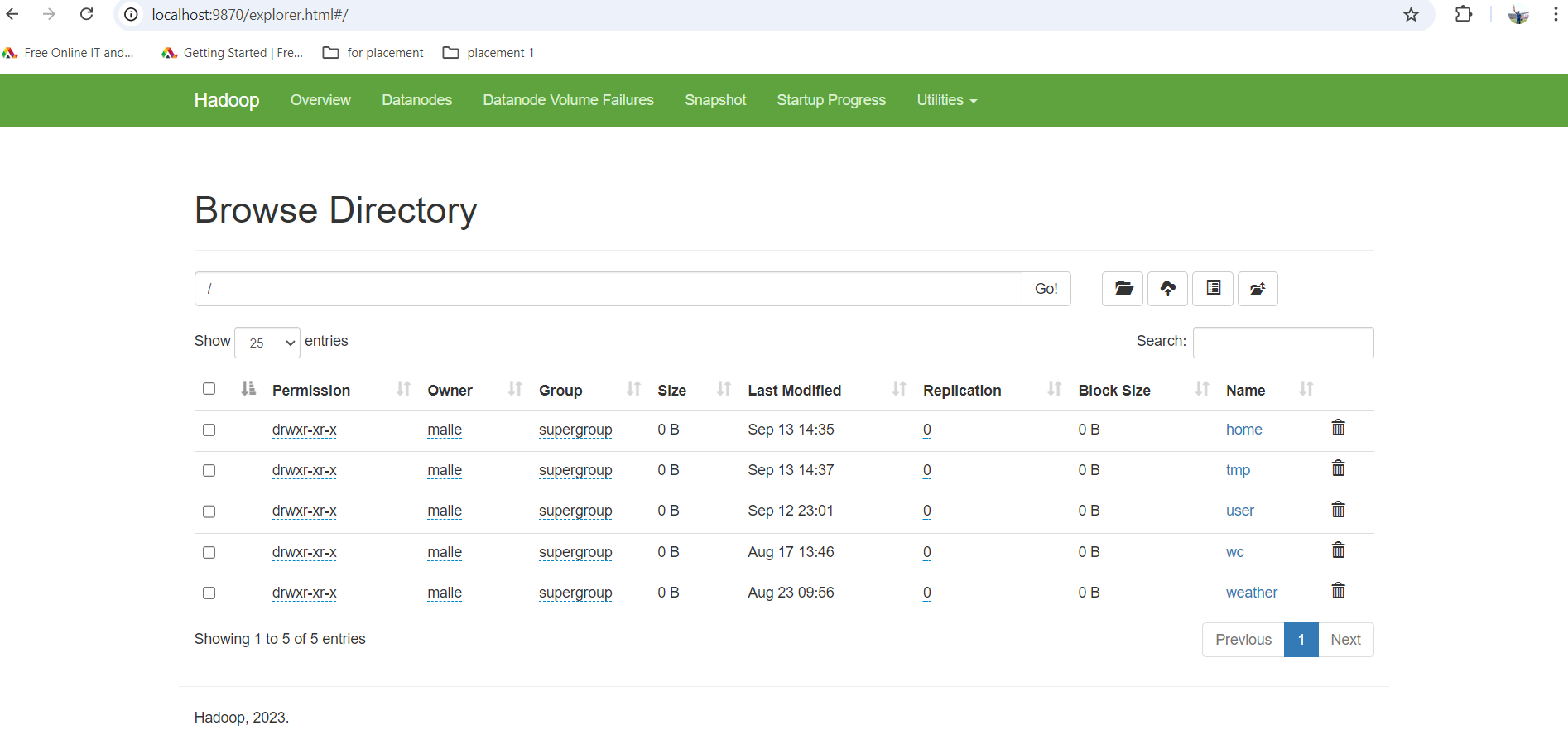
print('%s\t%s' % (prev\_word, prev\_count))

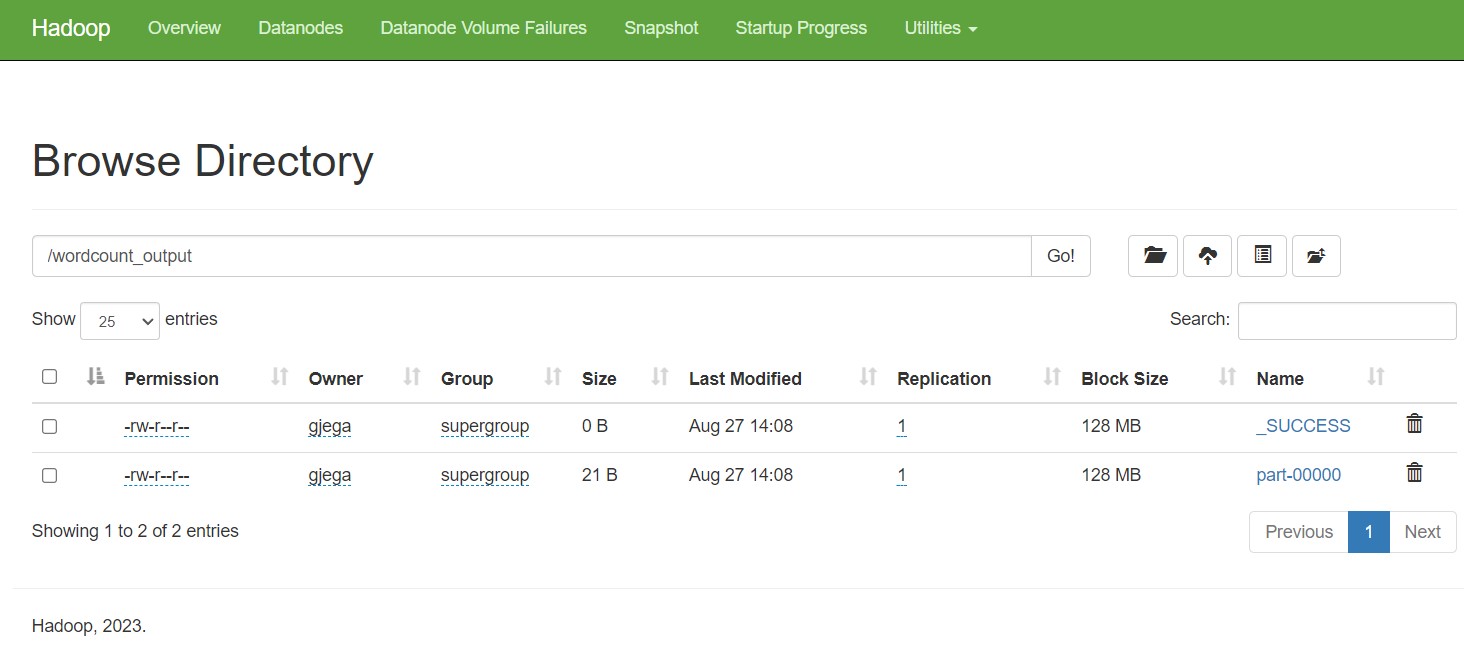
prev\_count = count prev\_word = word

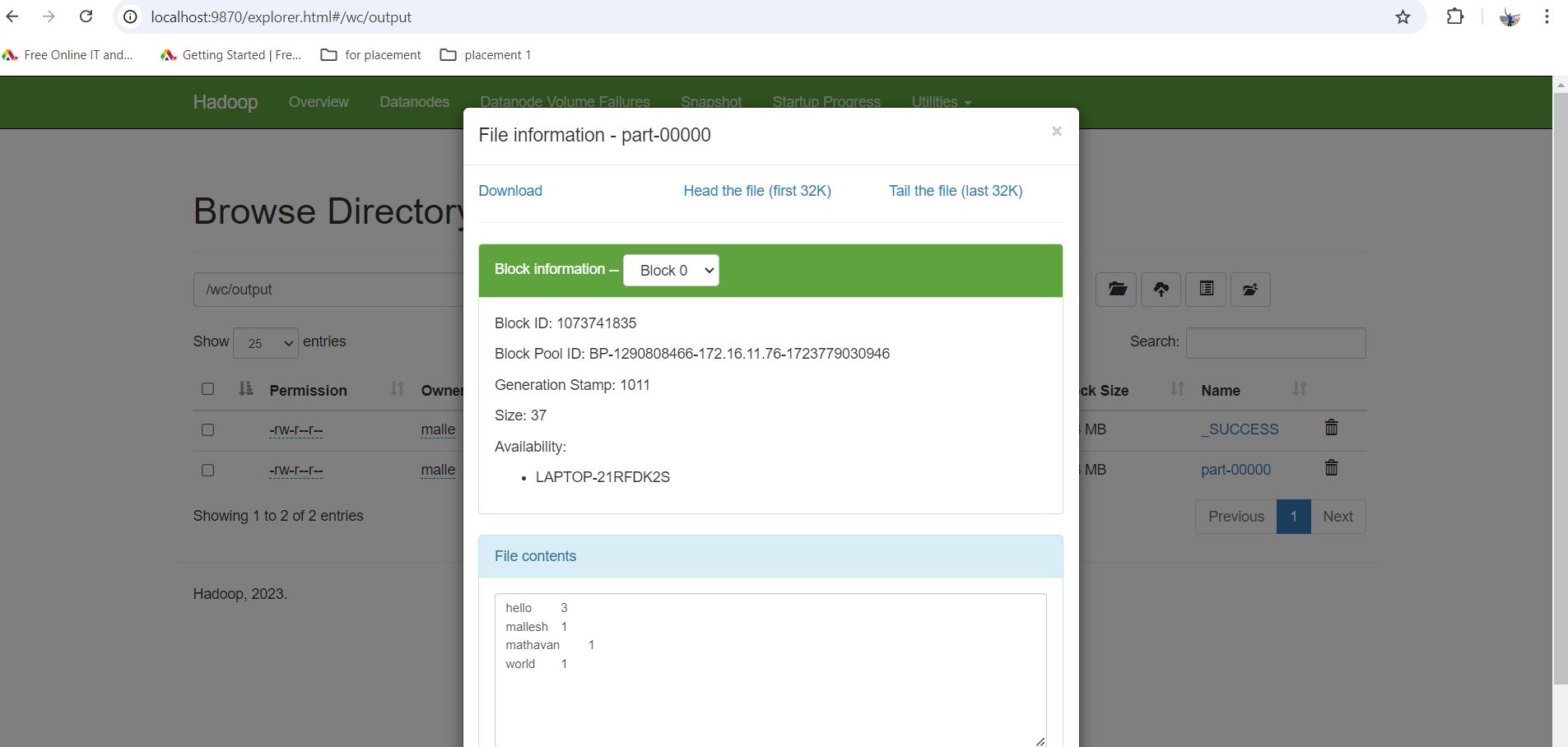
if prev\_word == word:

print('%s\t%s' % (prev\_word, prev\_count))

**OUTPUT:**





****

**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.