EX.NO. 2 **Roll No: 210701138**

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 and run as administrator and start the Hadoop by using the command:

```
C:\Windows\System32>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

C:\Windows\System32>jps
10068 NodeManager
30616 Jps
21308 DataNode
5612 ResourceManager
5836 NameNode
```

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

```
C:\Windows\System32>hdfs dfs -mkdir /words
```

3. Upload the input text file into the wordcount_ex2 directory using the command:

```
C:\Windows\System32>hdfs dfs -put C:\Users\Manoj\Desktop\word\input.txt /word
```

4. Create the mapper and reducer files.

5. To execute the files with Hadoop streaming run the following command:

C:\Windows\System32>hadoop jar "C:\hadoop\share\hadoop\tools\lib\hadoop-streaming-3.3.6.jar" ^-input /wordcount_ex2/word.txt ^-output /wordcount_op ^-mapper "python C:\Users\Manoj\Desktop\word\mapper.py" ^-reducer "python C:\Users\Manoj\Desktop\word\mapper.py"

MAPPER.PY

```
#!/usr/bin/env python
import sys
```

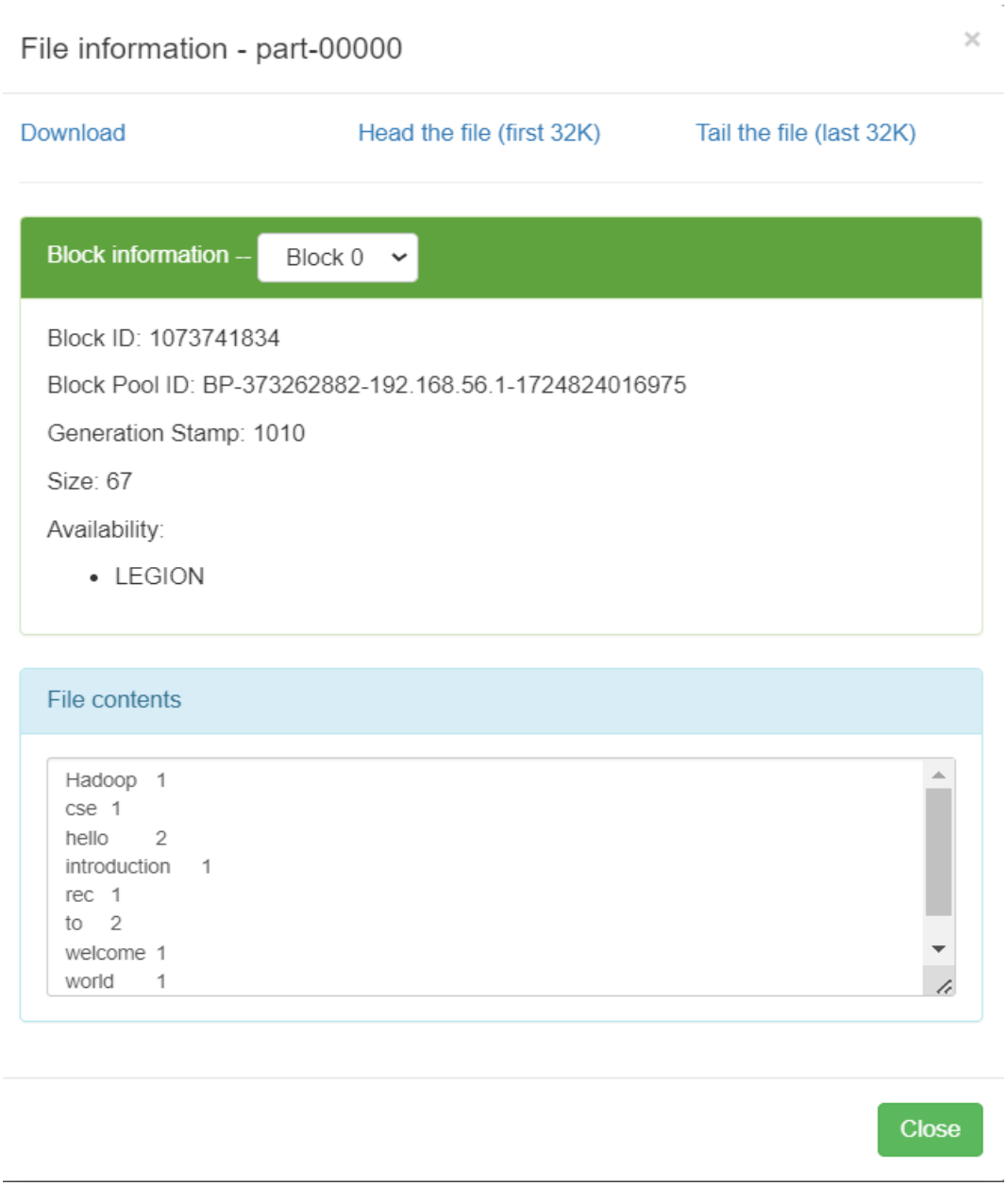
```
# Read lines from standard input
for line in sys.stdin:
    # Strip leading and trailing whitespaces
    line = line.strip()



    # Split the line into words
    words = line.split()

    # Output each word with a count of 1
    for word in words:
        print(f'{word}\t1')
```

REDUCER.PY

```
#!/usr/bin/env python
import sys
from collections import defaultdict

# Initialize a dictionary to store word counts
word_count = defaultdict(int)

# Read lines from standard input
for line in sys.stdin:
    # Strip leading and trailing whitespaces
    line = line.strip()

    # Split the line into word and count
    word, count = line.split('\t', 1)

    try:
        count = int(count)
    except ValueError:
        # If count is not an integer, skip this line
        continue

    # Add the count to the word's total
    word_count[word] += count

# Output each word and its total count
for word, count in word_count.items():
    print(f'{word}\t{count}')
```

EX.NO. 2 **Roll No: 210701138**

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.