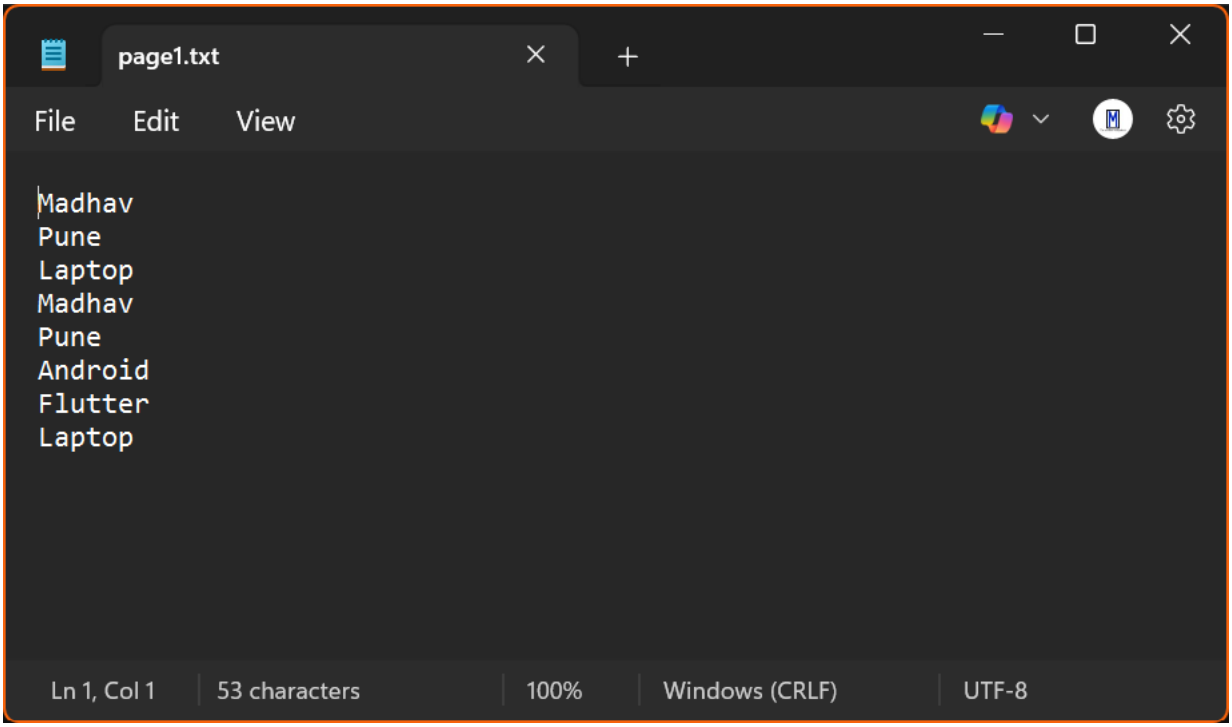
WordCount.java

```
package com.mapreduce.wc;
import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.util.GenericOptionsParser;

public class WordCount {

   public static void main(String[] args) throws Exception {
      Configuration c = new Configuration();
      String[] files = new GenericOptionsParser(c, args).getRemainingArgs();

      // Ensure correct input arguments
      if (files.length < 2) {
         System.err.println("Usage: WordCount <input path> <output path>");
         System.exit(-1);
      }

      Path input = new Path(files[0]);
```

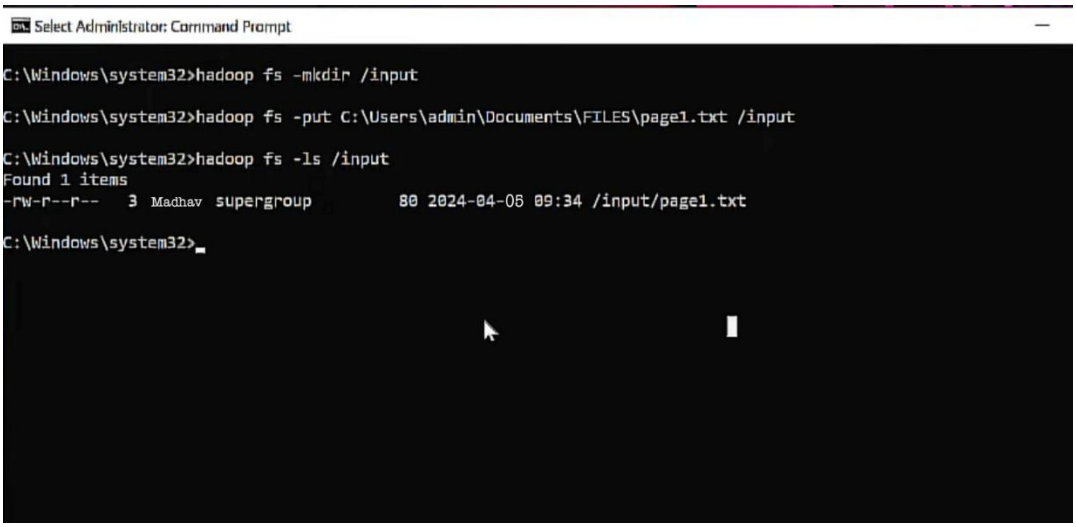
```
        Path output = new Path(files[1]);

        Job j = Job.getInstance(c, "wordcount");
        j.setJarByClass(WordCount.class);
        j.setMapperClass(MapForWordCount.class);
        j.setReducerClass(ReduceForWordCount.class);
        j.setOutputKeyClass(Text.class);
        j.setOutputValueClass(IntWritable.class);

        FileInputFormat.addInputPath(j, input);
        FileOutputFormat.setOutputPath(j, output);

        System.exit(j.waitForCompletion(true) ? 0 : 1);
    }

    // Mapper Class
    public static class MapForWordCount extends Mapper<LongWritable, Text, Text,
IntWritable> {
        private final static IntWritable one = new IntWritable(1);
        private Text wordText = new Text();

        public void map(LongWritable key, Text value, Context con) throws IOException,
InterruptedException {
            String line = value.toString().trim();
            String[] words = line.split("\\s+"); // Handles multiple spaces

            for (String word : words) {
                if (!word.isEmpty()) { // Avoid empty strings
                    wordText.set(word.trim().toUpperCase());
```

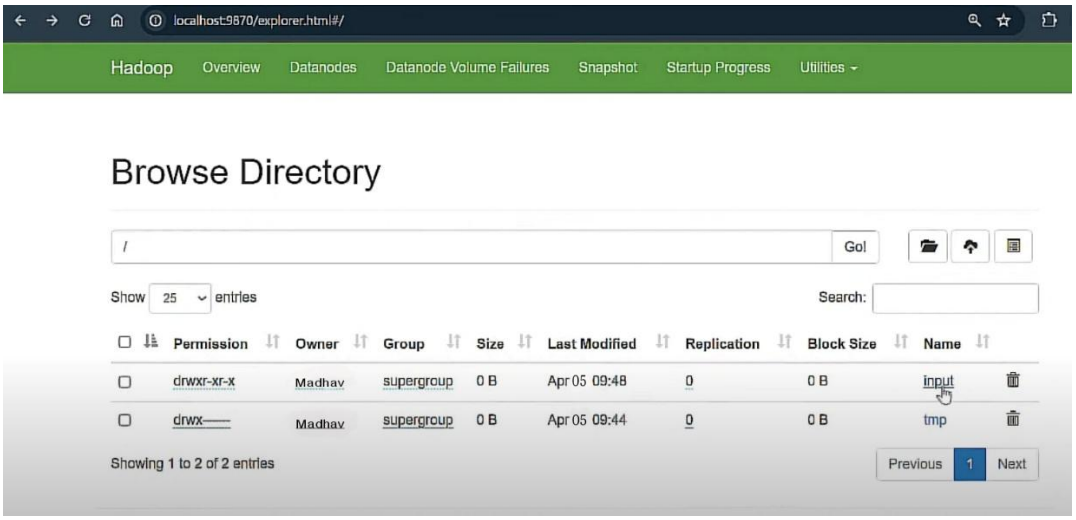
```
                con.write(wordText, one);
            }
        }
    }

    // Reducer Class
    public static class ReduceForWordCount extends Reducer<Text, IntWritable, Text, IntWritable> {
        public void reduce(Text word, Iterable<IntWritable> values, Context con) throws IOException, InterruptedException {
            int sum = 0;
            for (IntWritable value : values) {
                sum += value.get();
            }
            con.write(word, new IntWritable(sum));
        }
    }
}
```

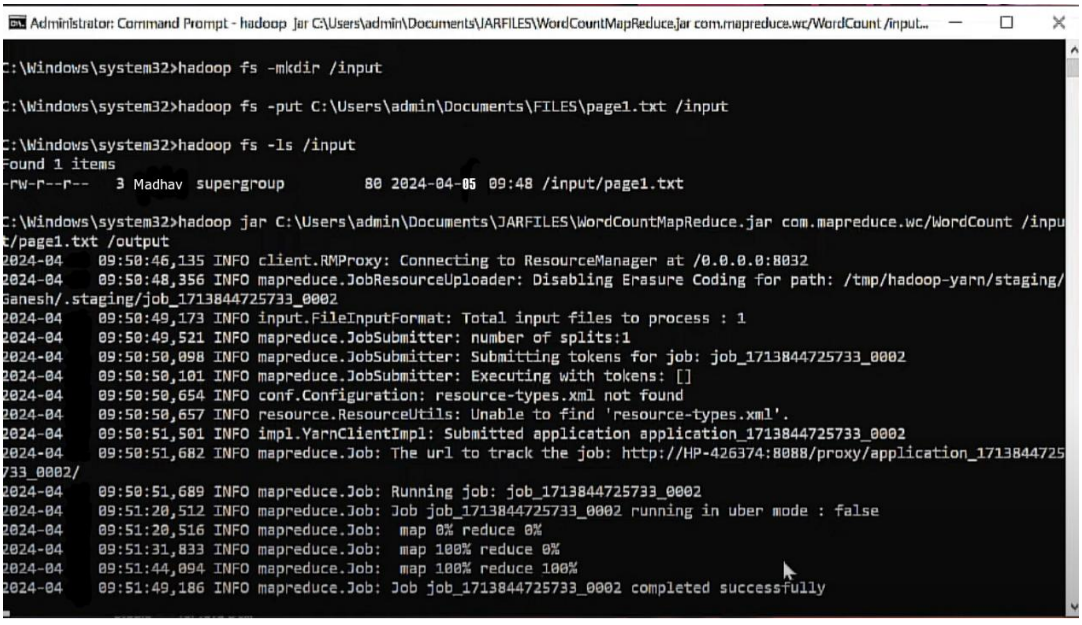
OUTPUTS:

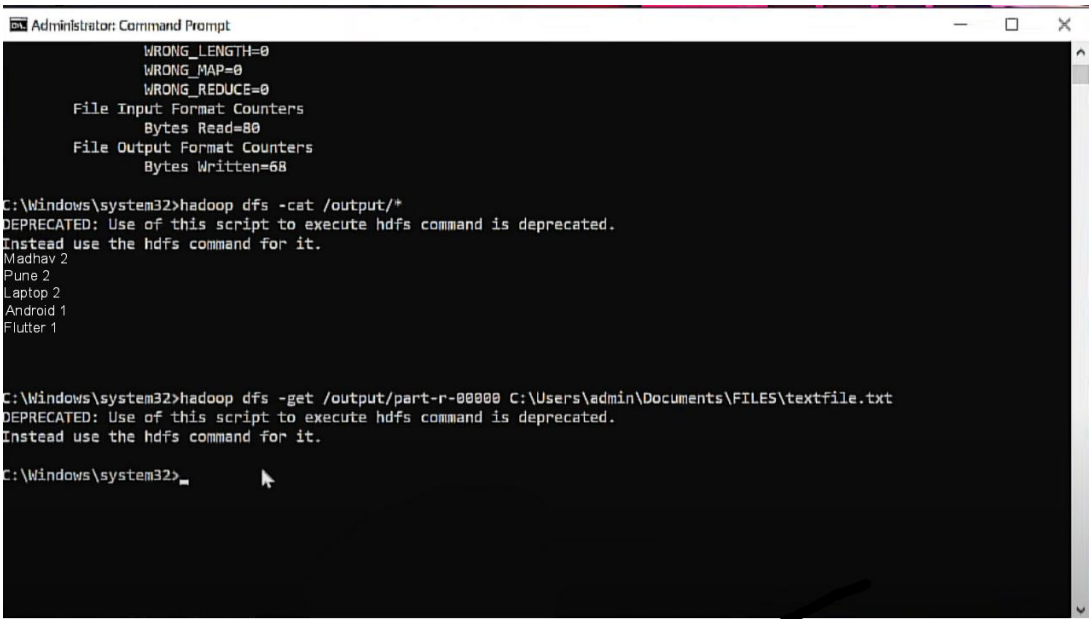


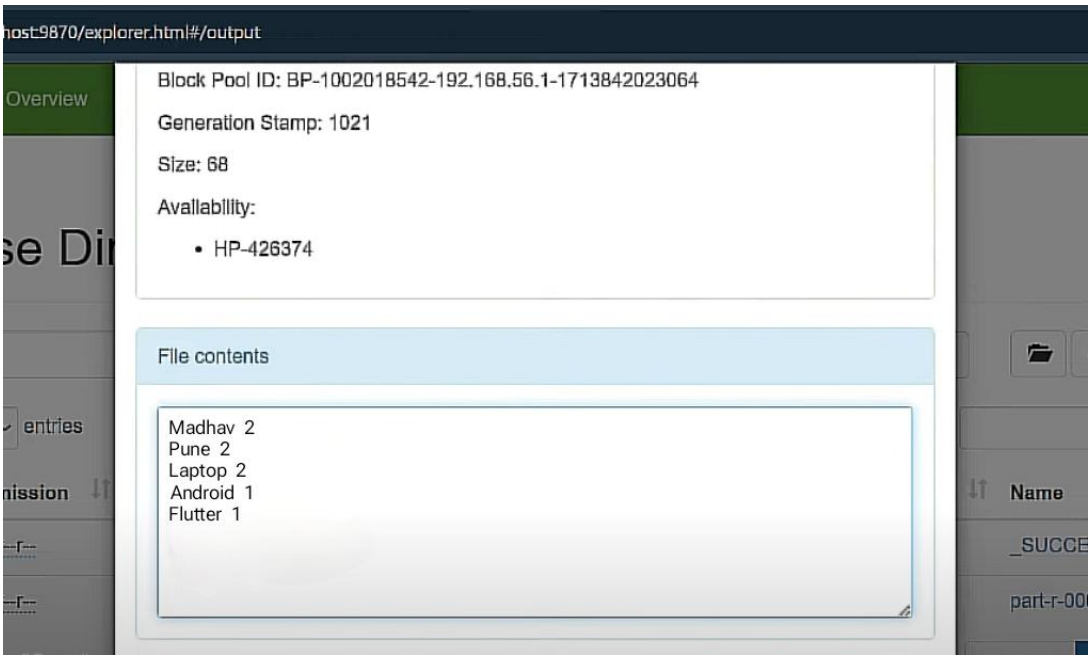

```
Select Administrator: Command Prompt
2024-04-22 09:24:28,157 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at DESKTOP-PL1EI1I/192.168.137.1
************************************************************/

C:\Windows\system32>cd \

C:\>cd hadoop

C:\hadoop>cd sbin

C:\hadoop\sbin>start-dfs.cmd

C:\hadoop\sbin>jps
13632 NameNode
6448 DataNode
12360 Jps

C:\hadoop\sbin>start-yarn.cmd
starting yarn daemons

C:\hadoop\sbin>jps
13632 NameNode
2496 ResourceManager
6448 DataNode
3604 NodeManager
16120 Jps

C:\hadoop\sbin>
```

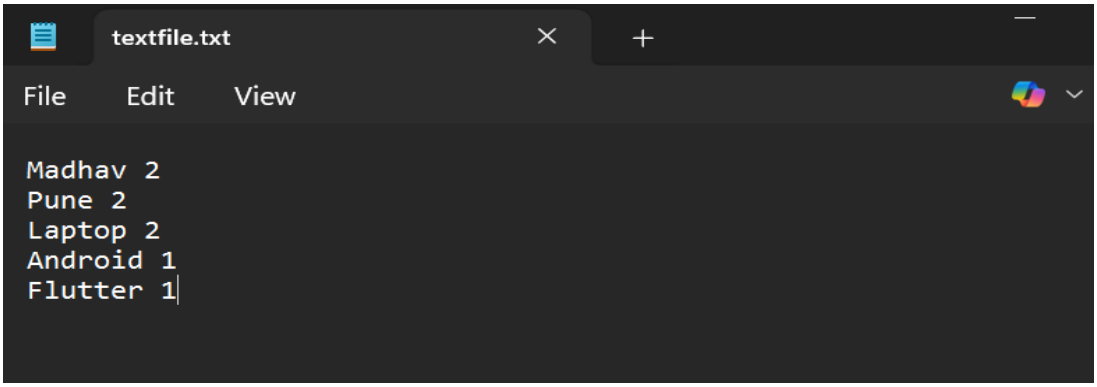

Select Administrator: Command Prompt
C:\Windows\system32>hadoop fs -mkdir /input
C:\Windows\system32>hadoop fs -put C:\Users\admin\Documents\FILES\page1.txt /input
C:\Windows\system32>hadoop fs -ls /input
Found 1 items
-rw-r--r-- 3 Madhav supergroup 80 2024-04-05 09:34 /input/page1.txt
C:\Windows\system32>


localhost:9870/explorer.html#/
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Browse Directory
/
Go!
Show 25 entries
Search:
Permission
Owner
Group
Size
Last Modified
Replication
Block Size
Name
drwxr-xr-x
Madhav
supergroup
0 B
Apr 05 09:48
0
0 B
input
drwx——
Madhav
supergroup
0 B
Apr 05 09:44
0
0 B
tmp
Showing 1 to 2 of 2 entries
Previous
1
Next


Administrator: Command Prompt - hadoop jar C:\Users\admin\Documents\JARFILES\WordCountMapReduce.jar com.mapreduce.wc/WordCount /input...
C:\Windows\system32>hadoop fs -mkdir /input
C:\Windows\system32>hadoop fs -put C:\Users\admin\Documents\FILES\page1.txt /input
C:\Windows\system32>hadoop fs -ls /input
Found 1 items
-rw-r--r-- 3 Madhav supergroup 80 2024-04-05 09:48 /input/page1.txt
C:\Windows\system32>hadoop jar C:\Users\admin\Documents\JARFILES\WordCountMapReduce.jar com.mapreduce.wc/WordCount /input/page1.txt /output
2024-04 09:50:46,135 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2024-04 09:50:48,356 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/Ganesh/.staging/job_1713844725733_0002
2024-04 09:50:49,173 INFO input.FileInputFormat: Total input files to process : 1
2024-04 09:50:49,521 INFO mapreduce.JobSubmitter: number of splits:1
2024-04 09:50:50,098 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1713844725733_0002
2024-04 09:50:50,101 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-04 09:50:50,654 INFO conf.Configuration: resource-types.xml not found
2024-04 09:50:50,657 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-04 09:50:51,501 INFO impl.YarnClientImpl: Submitted application application_1713844725733_0002
2024-04 09:50:51,682 INFO mapreduce.Job: The url to track the job: http://HP-426374:8088/proxy/application_1713844725733_0002/
2024-04 09:50:51,689 INFO mapreduce.Job: Running job: job_1713844725733_0002
2024-04 09:51:20,512 INFO mapreduce.Job: Job job_1713844725733_0002 running in uber mode : false
2024-04 09:51:20,516 INFO mapreduce.Job: map 0% reduce 0%
2024-04 09:51:31,833 INFO mapreduce.Job: map 100% reduce 0%
2024-04 09:51:44,094 INFO mapreduce.Job: map 100% reduce 100%
2024-04 09:51:49,186 INFO mapreduce.Job: Job job_1713844725733_0002 completed successfully


Administrator: Command Prompt
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=80
File Output Format Counters
Bytes Written=68
C:\Windows\system32>hadoop dfs -cat /output/*
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
Madhav 2
Pune 2
Laptop 2
Android 1
Flutter 1
C:\Windows\system32>hadoop dfs -get /output/part-r-00000 C:\Users\admin\Documents\FILES\textfile.txt
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
C:\Windows\system32>


host:9870/explorer.html#/output
Overview
Block Pool ID: BP-1002018542-192.168.56.1-1713842023064
Generation Stamp: 1021
Size: 68
Availability:
HP-426374
File contents
Madhav 2
Pune 2
Laptop 2
Android 1
Flutter 1
entries
Name


textfile.txt
File
Edit
View
Madhav 2
Pune 2
Laptop 2
Android 1
Flutter 1