Practical No. 02
Map Reduce using Hadoop

Location :


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

C:\Windows\System32>hdfs dfs -put "G:\hadoop-3.3.0\data.txt" /input

C:\Windows\System32>_
```
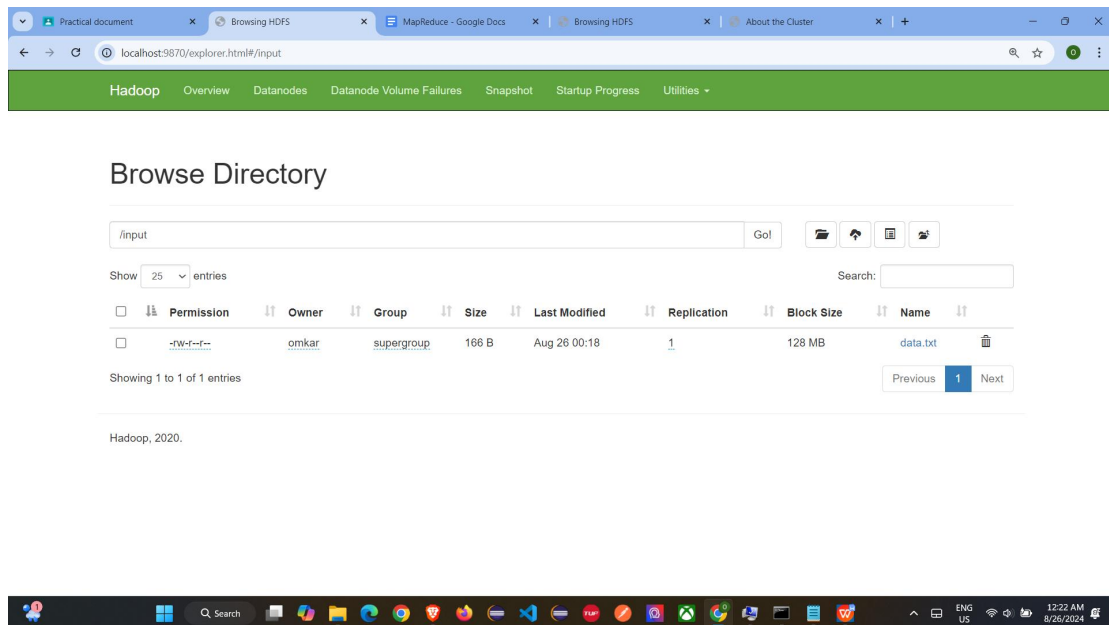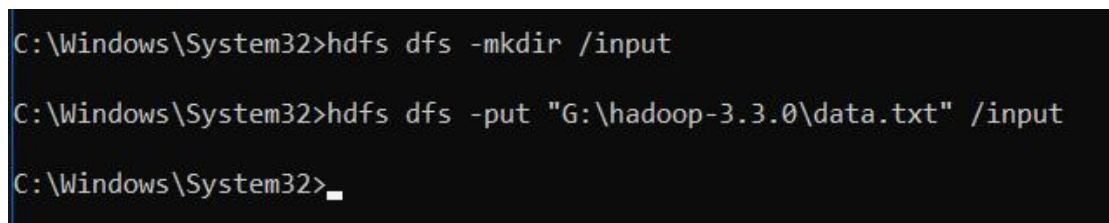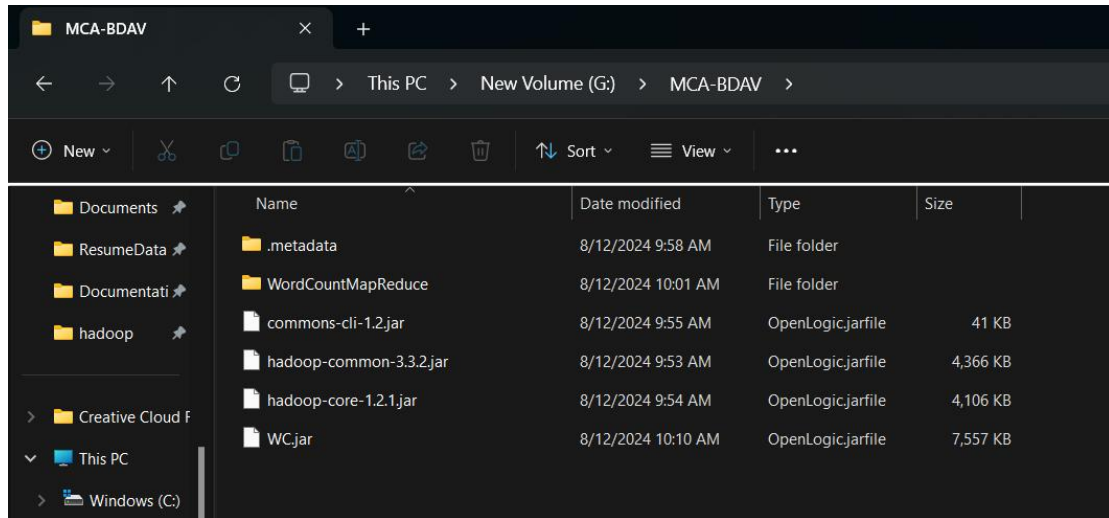
WordCountMapReduce.java

```java
package trial;

import java.io.IOException;
import java.util.StringTokenizer;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
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;

public class WordCountDemo {

  public static class TokenizerMapper
       extends Mapper<Object, Text, Text, IntWritable>{
    private final static IntWritable one = new IntWritable(1);
    private Text word = new Text();
    public void map(Object key, Text value, Context context) throws IOException, InterruptedException
{
      StringTokenizer itr = new StringTokenizer(value.toString());
      while (itr.hasMoreTokens()) {
       word.set(itr.nextToken());
       context.write(word, one);
      }
    }
  }

  public static class IntSumReducer
       extends Reducer<Text,IntWritable,Text,IntWritable> {
    private IntWritable result = new IntWritable();

    public void reduce(Text key, Iterable<IntWritable> values,
                Context context   ) throws IOException, InterruptedException {
      int sum = 0;
      for (IntWritable val : values) {
       sum += val.get();
      }
      result.set(sum);
      context.write(key, result);
    }
  }

  public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, "word count");
    job.setJarByClass(WordCountDemo.class);
    job.setMapperClass(TokenizerMapper.class);
    job.setCombinerClass(IntSumReducer.class);
    job.setReducerClass(IntSumReducer.class);
    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(IntWritable.class);
    FileInputFormat.addInputPath(job, new Path(args[0]));
```

```
        FileOutputFormat.setOutputPath(job, new Path(args[1]));
        System.exit(job.waitForCompletion(true) ? 0 : 1);
    }
}
```

WC.jar

```
C:\Windows\System32>hdfs dfs -cat /output/part-r-00000
Yyuyhcnhdhchtvfhfhfhfhfhrh          1


            Reduce output records=0
            Spilled Records=0
            Shuffled Maps =0
            Failed Shuffles=0
            Merged Map outputs=0
            GC time elapsed (ms)=0
            Total committed heap usage (bytes)=228065280
       Shuffle Errors
            BAD_ID=0
            CONNECTION=0
            IO_ERROR=0
            WRONG_LENGTH=0
            WRONG_MAP=0
            WRONG_REDUCE=0
       File Output Format Counters
            Bytes Written=0
024-08-12 10:15:49,020 INFO mapred.LocalJobRunner: Finishing task: attempt_local1309731849_0001_r_000000_0
024-08-12 10:15:49,021 INFO mapred.LocalJobRunner: reduce task executor complete.
024-08-12 10:15:49,623 INFO mapreduce.Job: Job job_local1309731849_0001 running in uber mode : false
024-08-12 10:15:49,627 INFO mapreduce.Job:  map 0% reduce 100%
024-08-12 10:15:49,631 INFO mapreduce.Job: Job job_local1309731849_0001 completed successfully
024-08-12 10:15:49,669 INFO mapreduce.Job: Counters: 30
       File System Counters
            FILE: Number of bytes read=7737941
            FILE: Number of bytes written=8413056
            FILE: Number of read operations=0
            FILE: Number of large read operations=0
            FILE: Number of write operations=0
C:\Windows\System32>hadoop jar "G:/MCA-BDAV/WC.jar" /input2 /output2
2024-08-12 10:15:47,931 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-08-12 10:15:47,985 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-08-12 10:15:47,986 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-08-12 10:15:48,265 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement
the Tool interface and execute your application with ToolRunner to remedy this.
2024-08-12 10:15:48,366 INFO input.FileInputFormat: Total input files to process : 0
2024-08-12 10:15:48,383 INFO mapreduce.JobSubmitter: number of splits:0
2024-08-12 10:15:48,445 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1309731849_0001
2024-08-12 10:15:48,445 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-08-12 10:15:48,606 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2024-08-12 10:15:48,607 INFO mapreduce.Job: Running job: job_local1309731849_0001
2024-08-12 10:15:48,610 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2024-08-12 10:15:48,619 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2024-08-12 10:15:48,619 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under outpu

C:\Windows\System32>hdfs dfsadmin -safemode leave
Safe mode is OFF


C:\Windows\System32>hdfs dfs -mkdir /input
mkdir: `/input': File exists


C:\Windows\System32>hdfs dfs -mkdir /input2


C:\Windows\System32>hdfs dfs -put "G:/hadoop-3.3.0/data.txt" /input2
put: `/hadoop-3.3.0/data.txt': No such file or directory


C:\Windows\System32>hdfs dfs -put "G:/hadoop-3.3.0/data.txt" /input2
put: `/hadoop-3.3.0/data.txt': No such file or directory

            Bytes written=100
2024-08-12 10:27:01,934 INFO mapred.LocalJobRunner: Finishing task: attempt_local1010835030_0001_r_000000_0
2024-08-12 10:27:01,935 INFO mapred.LocalJobRunner: reduce task executor complete.
2024-08-12 10:27:02,821 INFO mapreduce.Job:  map 100% reduce 100%
2024-08-12 10:27:02,822 INFO mapreduce.Job: Job job_local1010835030_0001 completed successfully
2024-08-12 10:27:02,831 INFO mapreduce.Job: Counters: 36
       File System Counters
            FILE: Number of bytes read=15476710
            FILE: Number of bytes written=16827194
            FILE: Number of read operations=0
            FILE: Number of large read operations=0
            FILE: Number of write operations=0
```

```
        at org.apache.hadoop.util.RunJar.main(RunJar.java:236)

C:\Windows\System32>hadoop jar "G:/MCA-BDAV/WC.jar" /input2 /output3
2024-08-12 10:26:59,822 INFO impl.MetricsConfig: Loaded properties from
2024-08-12 10:26:59,891 INFO impl.MetricsSystemImpl: Scheduled Metric s
2024-08-12 10:26:59 892 INFO impl MetricsSystemImpl: JobTracker metrics

C:\Windows\System32>hdfs dfs -put "G:\hadoop-3.3.0\spidy.txt" /input2


C:\Windows\System32>hdfs dfs -cat /output3/part-r-00000
Great   2
Ho      2
Jai     2
Man.    1
Onkar   1
Power   1
Responsibilities.       1
Spider  2
Text    1
This    2
Who     1
comes   1
day     1
demons  1
file.   2
great   1
is      3
is!     1
it      1
man?    1
the     1
with    1

C:\Windows\System32>
```