

Name : **Kartavya Mandani**

Roll No.: **20BCE120**

Course: **2CS702 Big Data Analytics**

Practical No.: **5 (MapReduce to find sum of given values)**

Steps:

- 1.Compilation of java file.
- 2.Convert class file to jar file.
- 3.Start Hadoop
- 4.Make directory in hdfs
- 5.Upload input file to the hdfs
- 6.input file format7.run Jar file command
- 8.Analysis the running Procedure of mapper and reducer
- 9.View the Output

```
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 Max{
    public static class TokenizerMapper
    extends Mapper<Object, Text, Text, IntWritable>{
        private static IntWritable num;
        private Text one = new Text("1");
        public void map(Object key, Text value, Context context) throws
        IOException,
        InterruptedException
        {
            StringTokenizer itr = new StringTokenizer(value.toString());
```

```

while (itr.hasMoreTokens())
{
    num = new IntWritable(Integer.parseInt(itr.nextToken()));
    context.write(one, num);
}
}
}

public static class IntSumReducer
extends Reducer<Text,IntWritable,Text,IntWritable> {
    public void reduce(Text key, Iterable<IntWritable>
    values,Context context)
    throws IOException, InterruptedException {
        int sum = 0;for (IntWritable val : values) {
        int temp = val.get();
        sum += temp;
        }
        context.write(new Text("Sum : "), new IntWritable(sum));
    }
}

public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, "word count");
    job.setJarByClass(Max.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);
}

```

Output:

```
Administrator: Command Prompt - hadoop jar "C:\0.Nirma University\0.SEM-7\2.Big Data Analytics\MapReduce\WordCount\wordco...
drwxr-xr-x - VISHV supergroup 0 2023-08-25 09:54 /Demo
-rw-r--r-- 1 VISHV supergroup 55 2023-10-12 21:52 /input.txt
-rw-r--r-- 1 VISHV supergroup 14 2023-10-13 09:39 /input2.txt
drwxr-xr-x - VISHV supergroup 0 2023-09-28 22:37 /test
drwx----- VISHV supergroup 0 2023-09-28 22:45 /tmp
drwxr-xr-x - VISHV supergroup 0 2023-10-13 09:05 /user
drwxr-xr-x - VISHV supergroup 0 2023-08-25 09:10 /vishv

C:\hadoopsetup\hadoop-3.2.4\sbin>hadoop jar "C:\0.Nirma University\0.SEM-7\2.Big Data Analytics\MapReduce\WordCount\word
count.jar" /input2.txt output2
2023-10-13 09:44:34,099 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-10-13 09:44:34,527 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement
the Tool interface and execute your application with ToolRunner to remedy this.
2023-10-13 09:44:34,551 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/
VISHV/.staging/job_1697168112381_0003
2023-10-13 09:44:34,697 INFO input.FileInputFormat: Total input files to process : 1
2023-10-13 09:44:34,760 INFO mapreduce.JobSubmitter: number of splits:1
2023-10-13 09:44:34,846 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1697168112381_0003
2023-10-13 09:44:34,848 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-10-13 09:44:34,964 INFO conf.Configuration: resource-types.xml not found
2023-10-13 09:44:34,965 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-10-13 09:44:35,016 INFO impl.YarnClientImpl: Submitted application application_1697168112381_0003
2023-10-13 09:44:35,038 INFO mapreduce.Job: The url to track the job: http://LAPTOP-1G612DH6:8088/proxy/application_1697
168112381_0003/
2023-10-13 09:44:35,039 INFO mapreduce.Job: Running job: job_1697168112381_0003
2023-10-13 09:44:41,163 INFO mapreduce.Job: Job job_1697168112381_0003 running in uber mode : false
2023-10-13 09:44:41,166 INFO mapreduce.Job: map 0% reduce 0%
2023-10-13 09:44:45,261 INFO mapreduce.Job: map 100% reduce 0%
2023-10-13 09:44:50,345 INFO mapreduce.Job: map 100% reduce 100%
```

```
Administrator: Command Prompt

Reduce input records=4
Reduce output records=4
Spilled Records=8
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=61
CPU time spent (ms)=0
Physical memory (bytes) snapshot=528084992
Virtual memory (bytes) snapshot=779894784
Total committed heap usage (bytes)=393216000
Peak Map Physical memory (bytes)=316022784
Peak Map Virtual memory (bytes)=389685248
Peak Reduce Physical memory (bytes)=212062208
Peak Reduce Virtual memory (bytes)=390209536
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
WordCount2$TokenizerMapper$CountersEnum
INPUT_WORDS=4
File Input Format Counters
Bytes Read=14
File Output Format Counters
Bytes Written=20

\hadoopsetup\hadoop-3.2.4\sbin>
```

Output

```
C:\hadoopsetup\hadoop-3.2.4\sbin>hdfs dfs -ls output2
Found 2 items
-rw-r--r--  1 VISHV supergroup      0 2023-10-13 10:08 output2/_SUCCESS
-rw-r--r--  1 VISHV supergroup    11 2023-10-13 10:08 output2/part-r-00000
```

```
C:\hadoopsetup\hadoop-3.2.4\sbin>hdfs dfs -cat output2/part-r-00000
Sum : 550
```

THE END

:)