Aim: Design and implement the MapReduce program to find the sum of given values.

Output:

SumDriver.java

```java
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class SumDriver {

    public static void main(String[] args) throws Exception {
        Configuration conf = new Configuration();
        Job job = Job.getInstance(conf, "Sum Calculation");

        job.setJarByClass(SumDriver.class);
        job.setMapperClass(SumMapper.class);
        job.setReducerClass(SumReducer.class);

        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(LongWritable.class);

        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));

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

SumMapper.java

```java
import java.io.IOException;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.*;
```

```java
public class SumMapper extends Mapper<LongWritable, Text, Text, LongWritable>
{
    private Text outputKey = new Text("sum");
    private LongWritable outputValue = new LongWritable();

    @Override
    public void map(LongWritable key, Text value, Context context) throws IOException,
            InterruptedException {
        try {
            // Parse the input value as a long
            long inputValue = Long.parseLong(value.toString());

            // Set the output value as the parsed long
            outputValue.set(inputValue);

            // Emit the key-value pair
            context.write(outputKey, outputValue);
        } catch (NumberFormatException e) {
            // Handle invalid input (non-numeric values) if necessary
        }
    }
}
```

SumReducer.java

```java
import java.io.IOException;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.*;

public class SumReducer extends Reducer<Text, LongWritable, Text,
LongWritable> {
    private LongWritable result = new LongWritable();

    @Override
    public void reduce(Text key, Iterable<LongWritable> values, Context
context)
            throws IOException, InterruptedException {
        long sum = 0;

        // Calculate the sum of all values in the iterable
        for (LongWritable value : values) {
            sum += value.get();
        }

        // Set the result as the sum
        result.set(sum);
```
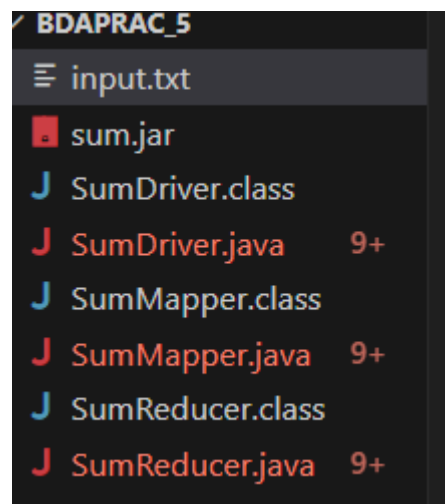
```
        // Emit the final result
        context.write(key, result);
    }
}
```

BDAPRAC_5
- input.txt
- sum.jar
- SumDriver.class
- SumDriver.java   9+
- SumMapper.class
- SumMapper.java   9+
- SumReducer.class
- SumReducer.java   9+

input.txt
```
1    1
2    2
3    25
4    55
5    4554
6    4
```

```
PS D:\bdaprac_5> C:\Progra~1\Java\jdk-1.8\bin\javac -classpath $(hadoop classpath) -d . SumDriver.java SumMapper.java SumReducer.java
PS D:\bdaprac_5> jar cvf sum.jar *.class
added manifest
adding: SumDriver.class(in = 1356) (out= 750)(deflated 44%)
adding: SumMapper.class(in = 1738) (out= 709)(deflated 59%)
adding: SumReducer.class(in = 1695) (out= 715)(deflated 57%)
PS D:\bdaprac_5> hdfs dfs -mkdir /input5
PS D:\bdaprac_5> hdfs dfs -put D:\bdaprac_5\input.txt /input5
2023-10-27 10:00:41,504 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
PS D:\bdaprac_5> hadoop jar sum.jar SumDriver /input5/input.txt /output5
2023-10-27 10:01:16,528 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-10-27 10:01:17,240 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with Too
lRunner to remedy this.
2023-10-27 10:01:17,263 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/Ayush/.staging/job_1698379320721_0002
2023-10-27 10:01:17,352 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-27 10:01:17,461 INFO input.FileInputFormat: Total input files to process : 1
2023-10-27 10:01:17,509 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-27 10:01:17,549 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-27 10:01:17,567 INFO mapreduce.JobSubmitter: number of splits:1
2023-10-27 10:01:17,709 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2023-10-27 10:01:17,733 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1698379320721_0002
2023-10-27 10:01:17,733 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-10-27 10:01:17,932 INFO conf.Configuration: resource-types.xml not found
2023-10-27 10:01:17,933 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-10-27 10:01:18,264 INFO impl.YarnClientImpl: Submitted application application_1698379320721_0002
2023-10-27 10:01:18,311 INFO mapreduce.Job: The url to track the job: http://LAPTOP-H8K2URUV:8088/proxy/application_1698379320721_0002/
2023-10-27 10:01:18,312 INFO mapreduce.Job: Running job: job_1698379320721_0002
2023-10-27 10:01:25,459 INFO mapreduce.Job: Job job_1698379320721_0002 running in uber mode : false
2023-10-27 10:01:25,460 INFO mapreduce.Job:  map 0% reduce 0%
2023-10-27 10:01:30,574 INFO mapreduce.Job:  map 100% reduce 0%
```

```
2023-10-27 10:01:35,635 INFO mapreduce.Job:  map 100% reduce 100%
2023-10-27 10:01:35,641 INFO mapreduce.Job: Job job_1698379320721_0002 completed successfully
2023-10-27 10:01:35,786 INFO mapreduce.Job: Counters: 50
        File System Counters
                FILE: Number of bytes read=90
                FILE: Number of bytes written=453689
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=123
                HDFS: Number of bytes written=9
                HDFS: Number of read operations=8
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=2838
                Total time spent by all reduces in occupied slots (ms)=2949
                Total time spent by all map tasks (ms)=2838
                Total time spent by all reduce tasks (ms)=2949
                Total vcore-milliseconds taken by all map tasks=2838
                Total vcore-milliseconds taken by all reduce tasks=2949
                Total megabyte-milliseconds taken by all map tasks=2906112
```

```
                Combine input records=0
                Combine output records=0
                Reduce input groups=1
                Reduce shuffle bytes=90
                Reduce input records=6
                Reduce output records=1
                Spilled Records=12
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=94
                CPU time spent (ms)=0
                Physical memory (bytes) snapshot=0
                Virtual memory (bytes) snapshot=0
                Total committed heap usage (bytes)=295698432
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=21
        File Output Format Counters
                Bytes Written=9
PS D:\bdaprac_5> hdfs dfs -cat /output5/part-r-00000
2023-10-27 10:02:35,294 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
sum     4641
PS D:\bdaprac_5>
```