Aayush Shah
19BCE245
17 October 2022

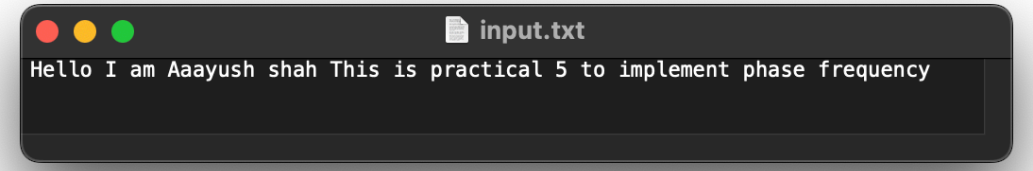
Big Data Analytics

Practical 5

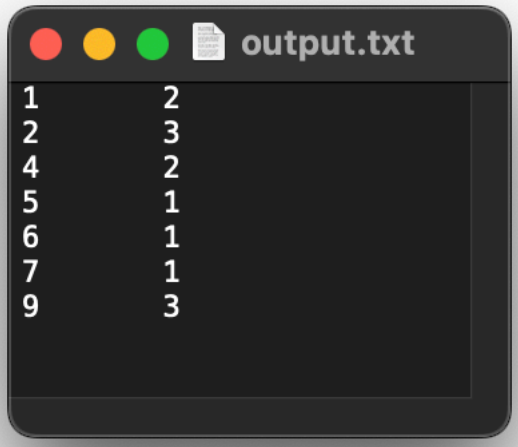
Aim

To implement phrase frequency using map-reduce programming.

Input



Output



Interpretation :

Words of length 1 : 2

Words of length 2 : 3 and so on…

Code

```
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 Prac5{
public static class TokenizerMapper
extends Mapper<Object, Text, Text, IntWritable>{
private final static IntWritable one = new
IntWritable(1); private Text word = new Text();
private int len;
private String length;
public void map(Object key, Text value, Context context
) throws IOException, InterruptedException {
StringTokenizer itr = new
StringTokenizer(value.toString());
while (itr.hasMoreTokens()) { len =
itr.nextToken().length();
length = Integer.toString(len); word.set(length);
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 length");
job.setJarByClass(Prac5.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);
} }
```

Conclusion

In this practical, I performed Phase frequency in hadoop in which we can know the count of word of specific lengths.