

Practical#5

Name: Saurin Anilkumar Prajapati

Roll No.: 19BCE239

Course Code and Name: 2CS702 Big Data Analytics

Batch: D1

AIM:

Apply MapReduce algorithms to find phrase frequency from given dataset.

□ Prepare a report to guide design of mapper and reducer.

Phase frequency Algorithm:

```
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 phrasefrequency {
    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 () ) {
                int len = itr . nextToken () . toString () . length () ;
                word . set (new Text ( Integer . toString ( len ) ) ) ;
                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 , " Phrase Frequency " ) ;
job . setJarByClass ( phrasefrequency . 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 ) ;
}
}

```

Thank You.