

Code:

```
package PackageDemo;

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

public class WordCount {
    public static void main(String[] args) throws Exception {
        Configuration conf = new Configuration();
        String[] files = new GenericOptionsParser(conf, args).getRemainingArgs();
        if (files.length < 2) {
            System.err.println("Usage: WordCount <input path> <output path>");
            System.exit(-1);
        }

        Path input = new Path(files[0]);
        Path output = new Path(files[1]);

        Job job = Job.getInstance(conf, "Word Count");
        job.setJarByClass(WordCount.class);
        job.setMapperClass(WordCountMapper.class);
        job.setReducerClass(WordCountReducer.class);

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

        FileInputFormat.addInputPath(job, input);
        FileOutputFormat.setOutputPath(job, output);

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

    public static class WordCountMapper extends Mapper<LongWritable, Text, Text, IntWritable> {
        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();

        public void map(LongWritable key, Text value, Context context) throws IOException,
        InterruptedException {
            String line = value.toString();
            String[] words = line.split("[\\s,;!?()]+"); // Splitting by spaces and punctuation
            for (String w : words) {
                if (!w.isEmpty()) {

```

```

        word.set(w);
        context.write(word, one);
    }
}
}
}

public static class WordCountReducer 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) {
            sum += val.get();
        }
        context.write(key, new IntWritable(sum));
    }
}
}

```

TXT file:

Hadoop is a powerful framework for distributed processing. MapReduce is used to process large datasets.

Big data technologies like Hadoop, Spark, and Flink are widely used in the industry. Data engineers rely on these tools.

The word count program is a basic yet essential Hadoop example. It demonstrates how MapReduce works.

Processing text files with Hadoop is efficient and scalable.

Machine learning and artificial intelligence are transforming the world. Companies use AI to analyze data.

Data-driven decision-making is key to business success.

The future of technology relies on big data and analytics. Cloud computing enables large-scale data processing.

Enterprises adopt data lakes and warehouses for structured and unstructured data.

Hadoop was inspired by Google's MapReduce and GFS (Google File System). It enables parallel processing.

Many companies like Facebook, Twitter, and Netflix use Hadoop for data storage and processing.

This text contains 500 words for testing. The word count should reflect the occurrences accurately. Repeat these lines until you reach approximately

Hadoop is a powerful framework for distributed processing. MapReduce is used to process large datasets.

Big data technologies like Hadoop, Spark, and Flink are widely used in the industry. Data engineers rely on these tools.

The word count program is a basic yet essential Hadoop example. It demonstrates how MapReduce works.

Processing text files with Hadoop is efficient and scalable.

Machine learning and artificial intelligence are transforming the world. Companies use AI to analyze data.

Data-driven decision-making is key to business success.

The future of technology relies on big data and analytics. Cloud computing enables large-scale data processing.

Enterprises adopt data lakes and warehouses for structured and unstructured data.

Hadoop was inspired by Google's MapReduce and GFS (Google File System). It enables parallel processing.

Many companies like Facebook, Twitter, and Netflix use Hadoop for data storage and processing.

This text contains 500 words for testing. The word count should reflect the occurrences accurately. Repeat these lines until you reach approximately

Hadoop is a powerful framework for distributed processing. MapReduce is used to process large datasets.

Big data technologies like Hadoop, Spark, and Flink are widely used in the industry. Data engineers rely on these tools.

The word count program is a basic yet essential Hadoop example. It demonstrates how MapReduce works.

Processing text files with Hadoop is efficient and scalable.

Machine learning and artificial intelligence are transforming the world. Companies use AI to analyze data.

Data-driven decision-making is key to business success.

The future of technology relies on big data and analytics. Cloud computing enables large-scale data processing.

Enterprises adopt data lakes and warehouses for structured and unstructured data.

Hadoop was inspired by Google's MapReduce and GFS (Google File System). It enables parallel processing.

Many companies like Facebook, Twitter, and Netflix use Hadoop for data storage and processing.

This text contains 500 words for testing. The word count should reflect the occurrences accurately. Repeat these lines until you reach approximately

Hadoop is a powerful framework for distributed processing. MapReduce is used to process large datasets.

Big data technologies like Hadoop, Spark, and Flink are widely used in the industry. Data engineers rely on these tools.

The word count program is a basic yet essential Hadoop example. It demonstrates how MapReduce works.

Processing text files with Hadoop is efficient and scalable.

Machine learning and artificial intelligence are transforming the world. Companies use AI to analyze data.

Data-driven decision-making is key to business success.

The future of technology relies on big data and analytics. Cloud computing enables large-scale data processing.

Enterprises adopt data lakes and warehouses for structured and unstructured data.

Hadoop was inspired by Google's MapReduce and GFS (Google File System). It enables parallel processing.

Many companies like Facebook, Twitter, and Netflix use Hadoop for data storage and processing.

This text contains 500 words for testing. The word count should reflect the occurrences accurately. Repeat these lines until you reach approximately

Hadoop is a powerful framework for distributed processing. MapReduce is used to process large datasets.

Big data technologies like Hadoop, Spark, and Flink are widely used in the industry. Data engineers rely on these tools.

The word count program is a basic yet essential Hadoop example. It demonstrates how MapReduce works.

Processing text files with Hadoop is efficient and scalable.

Machine learning and artificial intelligence are transforming the world. Companies use AI to analyze data.

Data-driven decision-making is key to business success.

The future of technology relies on big data and analytics. Cloud computing enables large-scale data processing.

Enterprises adopt data lakes and warehouses for structured and unstructured data.

Hadoop was inspired by Google's MapReduce and GFS (Google File System). It enables parallel processing.

Many companies like Facebook, Twitter, and Netflix use Hadoop for data storage and processing.

This text contains 500 words for testing. The word count should reflect the occurrences accurately. Repeat these lines until you reach approximately

#### OUTPUT:

A	15
ADOPT	5
AI	5
ANALYZE	5
ANALYTICS	5
AND	30
ARE	10
BASIC	5
BIG	10
BUSINESS	5

CLOUD	5	
COMPANIES	10	
COMPUTING	5	
CONTAINS	5	
COUNT	5	
DATA	25	
DATA-DRIVEN	5	
DATASETS	5	
DECISION-MAKING	5	5
DEMONSTRATES	5	5
DISTRIBUTED	5	
ENGINEERS	5	
ENTERPRISES	5	
ESSENTIAL	5	
FACEBOOK	5	
FILES	5	
FLINK	5	
FOR	20	
FRAMEWORK	5	5
GFS	5	
GOOGLE	5	
HADOOP	25	
INDUSTRY	5	
INSPIRED	5	
INTELLIGENCE	5	5
IS	30	
IT	15	
KEY	5	
LAKES	5	
LARGE	5	
LEARNING	5	5
LIKE	10	
MACHINE	5	
MAPREDUCE	15	
MANY	5	
MAPREDUCE	5	5
NETFLIX	5	
OF	20	
ON	5	
PARALLEL	5	
POWERFUL	5	
PROCESSING	15	
PROGRAM	5	
PROGRAMS	5	
REAL-TIME	5	
RELY	5	
SCALABLE	5	5
SPARK	5	
STORAGE	5	
STRUCTURED	5	5
SUCCESS	5	
TEXT	10	

THE	35	
THESE	5	
THEY	5	
TOOLS	5	
TOOLSET	5	
UNSTRUCTURED		5
USED	10	
WAREHOUSES		5
WAS	5	
WIDELY	5	
WITH	10	
WORD	10	
WORKS	5	
WORLD	5	