

ULLAAS M
INT19IS179

EXPERIMENT 3

```
WordCount.java
1 package ullaas;
2
3 import java.io.IOException;
4 import java.util.*;
5
6 import org.apache.hadoop.fs.Path;
7 import org.apache.hadoop.conf.*;
8 import org.apache.hadoop.io.*;
9 import org.apache.hadoop.mapred.*;
10 import org.apache.hadoop.util.*;
11
12 public class WordCount {
13
14     public static class Map extends MapReduceBase implements Mapper<LongWritable, Text, Text, IntWritable> {
15         private final static IntWritable one = new IntWritable(1);
16         private Text word = new Text();
17
18         public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output, Reporter reporter) throws IOException {
19             String line = value.toString();
20             StringTokenizer tokenizer = new StringTokenizer(line);
21             while (tokenizer.hasMoreTokens()) {
22                 word.set(tokenizer.nextToken());
23                 output.collect(word, one);
24             }
25         }
26     }
27
28     public static class Reduce extends MapReduceBase implements Reducer<Text, IntWritable, Text, IntWritable> {
29         public void reduce(Text key, Iterator<IntWritable> values, OutputCollector<Text, IntWritable> output, Reporter reporter) throws IOException {
30             int sum = 0;
31             while (values.hasNext()) {
32                 sum += values.next().get();
33             }
34         }
35     }
36 }
```

```
        output.collect(key, new IntWritable(sum));
    }
}

public static void main(String[] args) throws Exception {
    JobConf conf = new JobConf(WordCount.class);
    conf.setJobName("wordcount");

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

    conf.setMapperClass(Map.class);
    conf.setCombinerClass(Reduce.class);
    conf.setReducerClass(Reduce.class);

    conf.setInputFormat(TextInputFormat.class);
    conf.setOutputFormat(TextOutputFormat.class);

    FileInputFormat.setInputPaths(conf, new Path(args[0]));
    FileOutputFormat.setOutputPath(conf, new Path(args[1]));

    JobClient.runJob(conf);
}
```

```
cd $HADOOP_HOME
cd sbin
jps
start-all.sh
```

```

hadoop@admin1-HP-280-G4-MT-Business-PC:~/Desktop$ cd $HADOOP_HOME
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1$ cd sbin
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ jps
4051 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar
4694 Jps
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [admin1-HP-280-G4-MT-Business-PC]
Starting resourcemanager
Starting nodemanagers

```

```

hdfs dfs -mkdir -p ~ /input
hdfs dfs -appendToFile - ~/input/txt.txt
nmit
nitte meenakshi nitte

```

```

hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -mkdir -p ~ /input
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -appendToFile - ~/input/txt.txt
nmit
nitte meenakshi nitte

```

```

hadoop jar /home/hadoop/ullaas.jar ~/input ~/ullaas

```

```

hdfs dfs -cat ~/ullaas/part*

```

```

hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -cat ~/ullaas/part*
2022-06-18 12:35:03,618 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted
= false
meenakshi      1
nitte          2
nmit           1

```