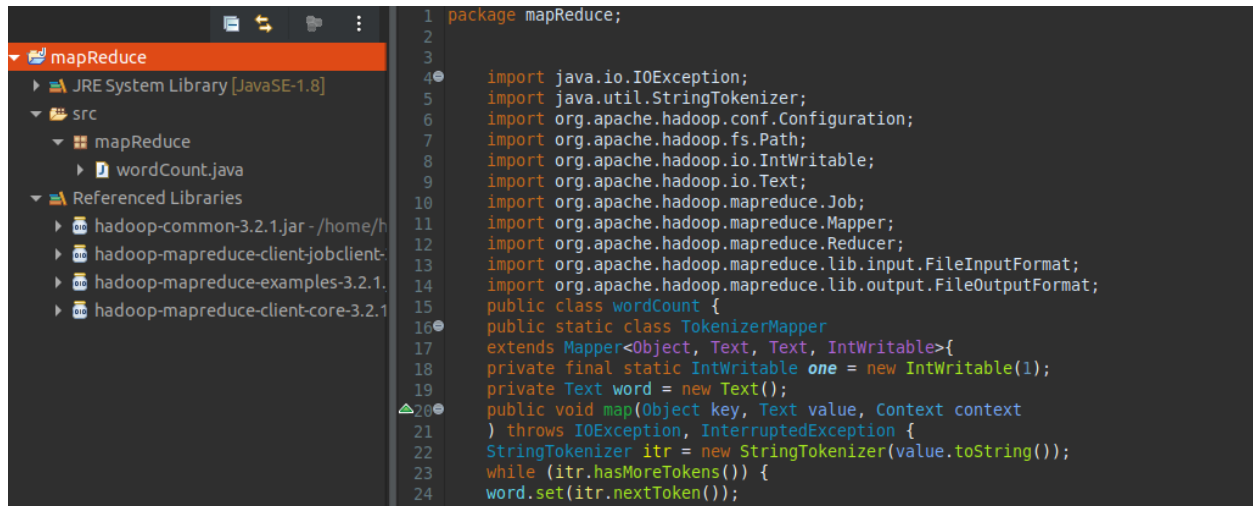# Map Reduce

NAME: T H LAKSHANA
USN: 1NT19IS174

```java
package mapReduce;


    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 wordCount {
    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()) {
    word.set(itr.nextToken());
```

```java
) 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 count");
job.setJarByClass(wordCount.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);
}
```

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~$ cd $HADOOP_HOME
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1$ cd sbin
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ ls
1nt19is097                      niki1nt19is097.txt       stop-all.sh
apache-hive-3.1.2-bin.tar.gz    p2.txt                   stop-balancer.sh
ch.txt                          p3.txt                   stop-dfs.cmd
day.txt                         refresh-namenodes.sh     stop-dfs.sh
distribute-exclude.sh           start-all.cmd            stop-secure-dns.sh
FederationStateStore            start-all.sh             stop-yarn.cmd
hadoop-daemon.sh                start-balancer.sh        stop-yarn.sh
hadoop-daemons.sh               start-dfs.cmd            workers.sh
httpfs.sh                       start-dfs.sh             x.txt
kms.sh                          start-secure-dns.sh      yarn-daemon.sh
mongodb-compass_1.26.1_amd64.deb start-yarn.cmd          yarn-daemons.sh
mr-jobhistory-daemon.sh         start-yarn.sh
mytext.txt                      stop-all.cmd
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ jps
5874 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar
6506 Jps
```

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hdoop 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
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -mkdir -p ~ /input
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -appendToFile - ~ /input/test.txt
appendToFile: stdin (-) must be the sole input argument when present
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -appendToFile  ~ /input/test.txt
appendToFile: /home/hdoop (Is a directory)
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -appendToFile - ~/input/test.txt
my name is lakshana my name is greeshma2022-06-17 09:26:55,650 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted
= false, remoteHostTrusted = false
```

```
    at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hadoop jar /home/hdoop/Desktop/lakshana.jar ~/input ~/out
2022-06-17 09:31:55,017 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-17 09:31:55,304 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and
 execute your application with ToolRunner to remedy this.
2022-06-17 09:31:55,343 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_165
5437936096_0001
2022-06-17 09:31:55,457 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-17 09:31:55,556 INFO input.FileInputFormat: Total input files to process : 1
2022-06-17 09:31:55,636 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-17 09:31:55,696 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-17 09:31:55,718 INFO mapreduce.JobSubmitter: number of splits:1
2022-06-17 09:31:55,826 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-17 09:31:56,252 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1655437936096_0001
2022-06-17 09:31:56,252 INFO mapreduce.JobSubmitter: Executing with tokens: []
```

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -cat ~/out/part*
2022-06-17 09:32:46,136 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
greeshma        1
is      2
lakshana        1
my      2
name    2
```