

20/4/25



PAGE:

DATE: / /

lab 7:

mapper code

// importing libraries

```

import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reporter;

```

```

public class WCMapper extends MapReduceBase implements
Mapper<LongWritable, Text, Text, IntWritable> {

```

// Map function

```

    public void map(LongWritable key, Text value,
                    OutputCollector<Text, IntWritable>
                    output, Reporter rep) throws
                    IOException
    {

```

```

        String line = value.toString();

```

```

        // Splitting the line on spaces
        for (String word : line.split(" ")) {

```

```

            if (word.length() > 0) {

```

```

                output.collect(new Text(word), new
                                IntWritable(1));
            }
        }
    }
}

```




```

import java.io.IOException;
import java.util.Iterator;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reducer;
import org.apache.hadoop.mapred.Reporter;

```

```

public class WReducer extends Mapper implements
    Reducer<Text, IntWritable, Text, IntWritable> {
    // Reduce Function

```

```

    public void reduce(Text key, Iterator<IntWritable>
        value, OutputCollector<Text, IntWritable> output,
        Reporter rep) throws IOException {

```

```

        int count = 0;
        while (value.hasNext()) {
            IntWritable i = value.next();
            count += i.get();
        }

```

```

        output.collect(key, new IntWritable(count));
    }
}

```




```

import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.FileInputFormat;
import org.apache.hadoop.mapred.FileOutputFormat;
import org.apache.hadoop.mapred.JobClient;
import org.apache.hadoop.mapred.JobConf;
import org.apache.hadoop.util.Tool;
import org.apache.hadoop.util.ToolRunner;

```

```

public class WCDriver extends Configuration
implements Tool {
    public int run(String args[]) throws IOException
    {

```

```

        if (args.length < 2)
        {

```

```

            System.out.println("Please give valid inputs");
            return -1;
        }

```

```

        JobConf conf = new JobConf(WCDriver.class);
        FileInputFormat.setInputPaths(conf, new Path(args[0]));
        FileOutputFormat.setOutputPath(conf, new Path(args[1]));
        conf.setReducerClass(WCReducer.class);
        conf.setMapperClass(WCMapper.class);
        conf.setMapOutputKeyClass(Text.class);
        conf.setMapOutputValueClass(IntWritable.class);
        JobClient.runJob(conf);
        return 0;
    }
}

```




```
public static void main(String args[]) throws Exception
```

```
int exitCode = ToolRunner.run(new WCDownloader(),  
                                args);
```

```
System.out.println(exitCode);
```

```
}
```

20/5/15

```
Activities Terminal May 20 14:47
hadoop@bmscscse-HP-Elite-Tower-600-G9-Desktop-PC: ~

hadoop@bmscscse-HP-Elite-Tower-600-G9-Desktop-PC: $ 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]
localhost: namenode is running as process 8499. Stop it first and ensure /tmp/hadoop-hadoop-namenode.pid file is empty before retry.
Starting datanodes
localhost: datanode is running as process 8673. Stop it first and ensure /tmp/hadoop-hadoop-datanode.pid file is empty before retry.
Starting secondary namenodes [bmscscse-HP-Elite-Tower-600-G9-Desktop-PC]
bmscscse-HP-Elite-Tower-600-G9-Desktop-PC: secondarynamenode is running as process 8959. Stop it first and ensure /tmp/hadoop-hadoop-secondarynamenode.pid file is empty before retry.
Starting resourcemanager
resourcemanager is running as process 9238. Stop it first and ensure /tmp/hadoop-hadoop-resourcemanager.pid file is empty before retry.
Starting nodemanagers
localhost: nodemanager is running as process 9399. Stop it first and ensure /tmp/hadoop-hadoop-nodemanager.pid file is empty before retry.
hadoop@bmscscse-HP-Elite-Tower-600-G9-Desktop-PC: $
nano /home/hadoop/hadoop/etc/hadoop/mapred-site.xml
hadoop@bmscscse-HP-Elite-Tower-600-G9-Desktop-PC: $ stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [bmscscse-HP-Elite-Tower-600-G9-Desktop-PC]
Stopping nodemanagers
Stopping resourcemanager
hadoop@bmscscse-HP-Elite-Tower-600-G9-Desktop-PC: $ 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 [bmscscse-HP-Elite-Tower-600-G9-Desktop-PC]
Starting resourcemanager
Starting nodemanagers
hadoop@bmscscse-HP-Elite-Tower-600-G9-Desktop-PC: $ jps
14785 DataNode
15187 SecondaryNameNode
15989 Jps
15386 ResourceManager
15741 NodeManager
6270 org.eclipse.equinox.launcher_1.6.1000.v20250227-1734.jar
14591 NameNode
hadoop@bmscscse-HP-Elite-Tower-600-G9-Desktop-PC: $ hadoop fs -ls /
Found 3 items
drwxr-xr-x - hadoop supergroup 0 2025-05-20 13:40 /folder1
drwxr-xr-x - hadoop supergroup 0 2025-05-20 13:40 /folder2
drwxr-xr-x - hadoop supergroup 0 2025-05-20 13:43 /tmp
hadoop@bmscscse-HP-Elite-Tower-600-G9-Desktop-PC: $ hadoop fs -mkdir /rgs
hadoop@bmscscse-HP-Elite-Tower-600-G9-Desktop-PC: $ hadoop fs -copyFromLocal /home/hadoop/Desktop/sample.txt /rgs/test.txt
hadoop@bmscscse-HP-Elite-Tower-600-G9-Desktop-PC: $ hadoop jar /home/hadoop/Desktop/wordcount.jar WordCount.WCDriver /rgs/test.txt /rgs/output
2025-05-20 14:45:00,274 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2025-05-20 14:45:00,315 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2025-05-20 14:45:00,315 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2025-05-20 14:45:00,321 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2025-05-20 14:45:00,384 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-05-20 14:45:00,436 INFO mapred.FileInputFormat: Total input files to process : 1
2025-05-20 14:45:00,469 INFO mapreduce.JobSubmitter: number of splits:1
```

```
Activities Terminal May 20 14:48 hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: ~

HDFS: Number of bytes written=86
HDFS: Number of read operations=15
HDFS: Number of large read operations=0
HDFS: Number of write operations=4
HDFS: Number of bytes read erasure-coded=0

Map-Reduce Framework
  Map input records=4
  Map output records=13
  Map output bytes=116
  Map output materialized bytes=148
  Input split bytes=86
  Combine input records=0
  Combine output records=0
  Reduce input groups=12
  Reduce shuffle bytes=148
  Reduce input records=13
  Reduce output records=12
  Spilled Records=26
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=0
  Total committed heap usage (bytes)=1375731712

Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0

File Input Format Counters
  Bytes Read=66
File Output Format Counters
  Bytes Written=86

0
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: $ hadoop fs -ls /output/
ls: '/output/': No such file or directory
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: $ hadoop fs -ls /rgs/output/
Found 2 items
-rw-r--r-- 1 hadoop supergroup      0 2025-05-20 14:45 /rgs/output/_SUCCESS
-rw-r--r-- 1 hadoop supergroup    86 2025-05-20 14:45 /rgs/output/part-00000
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC: $ hadoop fs -cat /rgs/output/part-00000
am      1
are     1
becz    1
executed 1
feeling 1
good    1
httt    1
how     1
ti       2
program 1
the     1
you     1
hadoop@bmscecse-HP-Elite-Tower-600-G9-Desktop-PC:~$ ss
```