Big Data Exp 1B : 1211061

Word Count:

Driver:

```
//WCDriver.java
package wc;

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.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class WCDriver {

	public static void main(String[] args) throws Exception {
		Configuration conf = new Configuration();
		Job job = Job.getInstance(conf, "JobName");
		job.setJarByClass(wc.WCDriver.class);
		job.setMapperClass(wc.WCMapper.class);

		job.setCombinerClass(wc.WCReducer.class);
		job.setReducerClass(wc.WCReducer.class);

		// TODO: specify output types
		job.setMapOutputKeyClass(Text.class);
		job.setMapOutputValueClass(IntWritable.class);
		job.setOutputKeyClass(Text.class);
		job.setOutputValueClass(IntWritable.class);

		// TODO: specify input and output DIRECTORIES (not files)
		FileInputFormat.setInputPaths(job, new
Path("/home/kjsce/Desktop/word_count_input"));
		FileOutputFormat.setOutputPath(job, new
Path("/home/kjsce/Desktop/word_count_output"));

		if (!job.waitForCompletion(true))
			return;
	}

}
```

Reducer:

```
//WCReducer.java
package wc;

import java.io.IOException;
import java.util.Iterator;

import org.apache.hadoop.io.FloatWritable;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

public class WCReducer
		extends
		Reducer<Text, org.apache.hadoop.io.IntWritable, Text,
org.apache.hadoop.io.IntWritable> {

	public void reduce(Text _key, Iterable<IntWritable> values, Context context)
			throws IOException, InterruptedException {
		// process values
		Iterator<IntWritable> iterator=values.iterator();
		/*for (IntWritable val : values) {

		}*/
		int m=0;
		while(iterator.hasNext())
		{
			m=m+iterator.next().get();
		}
		//System.out.println(m);
		context.write(_key, new IntWritable(m));
	}

}
```

Mapper:

```
//WCMapper.java
package wc;


import java.io.IOException;
import java.util.Arrays;

import org.apache.commons.lang.StringUtils;
import org.apache.hadoop.io.FloatWritable;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;
```

```
public class WCMapper extends
		Mapper<LongWritable, Text, Text, org.apache.hadoop.io.IntWritable> {

	public void map(LongWritable ikey, Text ivalue, Context context)
			throws IOException, InterruptedException {
		String line=ivalue.toString();
		String[] tokens=StringUtils.split(line,' ');
		System.out.println(Arrays.toString(tokens));
		int size=tokens.length,c=1;
		for(int i=0; i<size; i++)
		{
			context.write(new Text(tokens[i]), new IntWritable(c));
		}

	}

}
```

Word_count_input:

```
this is a box
this is a block
the block is on the box
or is it in the box
```

Word_count_output:

```
a	2
block	2
box	3
in	1
is	4
it	1
on	1
or	1
the	3
this	2
```