

LARGE SCALE DATA PROCESSING (CSE 3025)

WIN SEMESTER 2017

NAME: B.SHUBANKAR

REGNO: 15BCE1123

PROFESSOR: DR. Maheswari N

LAB-2

11 December 2017

MAPREDUCE

CODE:

```
import java.io.IOException;
import java.util.*;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.conf.*;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.*;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;

public class WordCount {
    public static class Map 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();
            StringTokenizer tokenizer = new StringTokenizer(line);
            while (tokenizer.hasMoreTokens()) {
                word.set(tokenizer.nextToken());
                context.write(word, one);
            }
        }
    }
}
```

```
public static class Reduce 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));
    }
}

public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    Job job = new Job(conf, "wordcount");
    job.setJarByClass(WordCount.class);
    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(IntWritable.class);
    job.setMapperClass(Map.class);
    job.setReducerClass(Reduce.class);
    job.setInputFormatClass(TextInputFormat.class);
    job.setOutputFormatClass(TextOutputFormat.class);
    FileInputFormat.addInputPath(job, new Path(args[0]));
    FileOutputFormat.setOutputPath(job, new Path(args[1]));
    job.waitForCompletion(true);
}
}
```

OUTPUT

Activities Vmplayer Mon 13:49
Hadoop_ponny - VMware Workstation 12 Player (Non-commercial use only)

File Virtual Machine Help

HDFS/user/out2/part-r-00000 - Mozilla Firefox

VIT University Mail - Sent Mail HDFS/user/out2/part-r-00000 notepad.pw /lsdplab1 | Fre... HDFS/user

localhost:50075/browseBlock.jsp?blockId=108592292006648244&blockSize=50&genstamp=1478&filename=%2Fuser%2Fout2%2Fpart-r-00000&datanodePort=50 Google

File: /user/out2/part-r-00000

Goto: /user/out2 go

[Go back to dir listing](#)
[Advanced view/download options](#)

```
a,j 1
ahello 1
aish 1
dog 2
hello 2
sam 1
sent2 1
```

[Download this file](#)
[Tail this file](#)

Chunk size to view (in bytes, up to file's DFS block size): 32768 Refresh

Total number of blocks: 1
108592292006648244: [127.0.0.1:50010](#)

[Go back to DFS home](#)

Local logs

To release input, press Ctrl+Alt

```
Activities Vmplayer Mon 13:50
Hadoop_ponny - VMware Workstation 12 Player (Non-commercial use only)
File Virtual Machine Help
ponny@ubuntu: ~/aish
17/12/11 13:48:29 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
17/12/11 13:48:29 INFO input.FileInputFormat: Total input paths to process : 1
17/12/11 13:48:29 INFO util.NativeCodeLoader: Loaded the native-hadoop library
17/12/11 13:48:29 WARN snappy.LoadSnappy: Snappy native library not loaded
17/12/11 13:48:29 INFO mapred.JobClient: Running job: job_201712111307_0002
17/12/11 13:48:30 INFO mapred.JobClient: map 0% reduce 0%
17/12/11 13:48:44 INFO mapred.JobClient: map 100% reduce 0%
17/12/11 13:48:59 INFO mapred.JobClient: map 100% reduce 100%

17/12/11 13:49:04 INFO mapred.JobClient: Job complete: job_201712111307_0002
17/12/11 13:49:04 INFO mapred.JobClient: Counters: 29
17/12/11 13:49:04 INFO mapred.JobClient: Job Counters
17/12/11 13:49:04 INFO mapred.JobClient:   Launched reduce tasks=1
17/12/11 13:49:04 INFO mapred.JobClient:   SLOTS_MILLIS_MAPS=10889
17/12/11 13:49:04 INFO mapred.JobClient:   Total time spent by all reduces waiting after reserving slots (ms)=0
17/12/11 13:49:04 INFO mapred.JobClient:   Total time spent by all maps waiting after reserving slots (ms)=0
17/12/11 13:49:04 INFO mapred.JobClient:   Launched map tasks=1
17/12/11 13:49:04 INFO mapred.JobClient:   Data-local map tasks=1
17/12/11 13:49:04 INFO mapred.JobClient:   SLOTS_MILLIS_REDUCES=12709
17/12/11 13:49:04 INFO mapred.JobClient: File Output Format Counters
17/12/11 13:49:04 INFO mapred.JobClient:   Bytes Written=50
17/12/11 13:49:04 INFO mapred.JobClient: FileSystemCounters
17/12/11 13:49:04 INFO mapred.JobClient:   FILE_BYTES_READ=106
17/12/11 13:49:04 INFO mapred.JobClient:   HDFS_BYTES_READ=160
17/12/11 13:49:04 INFO mapred.JobClient:   FILE_BYTES_WRITTEN=43267
17/12/11 13:49:04 INFO mapred.JobClient:   HDFS_BYTES_WRITTEN=50
17/12/11 13:49:04 INFO mapred.JobClient: File Input Format Counters
17/12/11 13:49:04 INFO mapred.JobClient:   Bytes Read=50
17/12/11 13:49:04 INFO mapred.JobClient: Map-Reduce Framework
17/12/11 13:49:04 INFO mapred.JobClient:   Map output materialized bytes=106
17/12/11 13:49:04 INFO mapred.JobClient:   Map input records=6
17/12/11 13:49:04 INFO mapred.JobClient:   Reduce shuffle bytes=0
17/12/11 13:49:04 INFO mapred.JobClient:   Spilled Records=18
17/12/11 13:49:04 INFO mapred.JobClient:   Map output bytes=82
17/12/11 13:49:04 INFO mapred.JobClient:   Total committed heap usage (bytes)=177016832
17/12/11 13:49:04 INFO mapred.JobClient:   CPU time spent (ms)=420
17/12/11 13:49:04 INFO mapred.JobClient:   Combine input records=0
17/12/11 13:49:04 INFO mapred.JobClient:   SPLIT_RAW_BYTES=110
17/12/11 13:49:04 INFO mapred.JobClient:   Reduce input records=9
17/12/11 13:49:04 INFO mapred.JobClient:   Reduce input groups=7
17/12/11 13:49:04 INFO mapred.JobClient:   Combine output records=0
17/12/11 13:49:04 INFO mapred.JobClient:   Physical memory (bytes) snapshot=188436480
17/12/11 13:49:04 INFO mapred.JobClient:   Reduce output records=7
17/12/11 13:49:04 INFO mapred.JobClient:   Virtual memory (bytes) snapshot=767225856
17/12/11 13:49:04 INFO mapred.JobClient:   Map output records=9
ponny@ubuntu:~/aish$
ponny@ubuntu:~/aish$ ls
aish.txt lab2.jar s2.txt
ponny@ubuntu:~/aish$
```

To release input, press Ctrl+Alt