

## ***Assignment No 2 – (Program 1) (WordCount)***

***RN: 33316***

***Batch : K11***

### ***Program :***

```
package PackageDemo;
import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
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;
import org.apache.hadoop.util.GenericOptionsParser;

public class WordCount {
public static void main(String [] args) throws Exception
{
Configuration c=new Configuration();
String[] files=new GenericOptionsParser(c,args).getRemainingArgs();
Path input=new Path(files[0]);
Path output=new Path(files[1]);
Job j=new Job(c,"wordcount");
j.setJarByClass(WordCount.class);
j.setMapperClass(MapForWordCount.class);
j.setReducerClass(ReduceForWordCount.class);
j.setOutputKeyClass(Text.class);
j.setOutputValueClass(IntWritable.class);
FileInputFormat.addInputPath(j, input);
FileOutputFormat.setOutputPath(j, output);
System.exit(j.waitForCompletion(true)?0:1);
}
public static class MapForWordCount extends Mapper<LongWritable, Text, Text, IntWritable>{
public void map(LongWritable key, Text value, Context con) throws IOException,
InterruptedException
{
String line = value.toString();
String[] words=line.split(" ");
for(String word: words )
{
Text outputKey = new Text(word.toUpperCase().trim());
IntWritable outputValue = new IntWritable(1);
con.write(outputKey, outputValue);
}
}
```

```
}  
}
```

```
public static class ReduceForWordCount extends Reducer<Text, IntWritable, Text, IntWritable>  
{  
    public void reduce(Text word, Iterable<IntWritable> values, Context con) throws IOException,  
        InterruptedException  
    {  
        int sum = 0;  
        for(IntWritable value : values)  
        {  
            sum += value.get();  
        }  
        con.write(word, new IntWritable(sum));  
    }  
}
```

### ***WordFile :***

***Bus, Car, bus, car, train, car, bus, car, train, bus, TRAIN,BUS, buS, caR, CAR, car, BUS,  
TRAIN, Sahil, Raj, Sahil ,Raj ,Sahil ,Raj***

### ***Output :***

***/cloudera@quickstart 33316/\$ hadoop fs -put wordfile wordFile5***

***/cloudera@quickstart 33316/\$ hadoop jar sample.jar PackageDemo.WordCount  
wordFile5 MRDir4***

25/01/21 22:23:42 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032  
25/01/21 22:23:43 INFO input.FileInputFormat: Total input paths to process : 1  
25/01/21 22:23:43 INFO mapreduce.JobSubmitter: number of splits:1  
25/01/21 22:23:43 INFO mapreduce.JobSubmitter: Submitting tokens for job:  
job\_1737525093676\_0002  
25/01/21 22:23:44 INFO impl.YarnClientImpl: Submitted application  
application\_1737525093676\_0002  
25/01/21 22:23:44 INFO mapreduce.Job: The url to track the job:  
http://quickstart.cloudera:8088/proxy/application\_1737525093676\_0002/  
25/01/21 22:23:44 INFO mapreduce.Job: Running job: job\_1737525093676\_0002  
25/01/21 22:23:52 INFO mapreduce.Job: Job job\_1737525093676\_0002 running in uber mode :  
false

***25/01/21 22:23:52 INFO mapreduce.Job: map 0% reduce 0%***

***25/01/21 22:23:59 INFO mapreduce.Job: map 100% reduce 0%***

***25/01/21 22:24:07 INFO mapreduce.Job: map 100% reduce 100%***

25/01/21 22:24:07 INFO mapreduce.Job: Job job\_1737525093676\_0002 completed successfully  
25/01/21 22:24:07 INFO mapreduce.Job: Counters: 49  
File System Counters

FILE: Number of bytes read=260  
FILE: Number of bytes written=221199  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=256  
HDFS: Number of bytes written=34  
HDFS: Number of read operations=6  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2

#### Job Counters

Launched map tasks=1  
Launched reduce tasks=1  
Data-local map tasks=1  
Total time spent by all maps in occupied slots (ms)=4676  
Total time spent by all reduces in occupied slots (ms)=5926  
Total time spent by all map tasks (ms)=4676  
Total time spent by all reduce tasks (ms)=5926  
Total vcore-seconds taken by all map tasks=4676  
Total vcore-seconds taken by all reduce tasks=5926  
Total megabyte-seconds taken by all map tasks=4788224  
Total megabyte-seconds taken by all reduce tasks=6068224

#### Map-Reduce Framework

Map input records=1  
Map output records=24  
Map output bytes=206  
Map output materialized bytes=260  
Input split bytes=120  
Combine input records=0  
Combine output records=0  
Reduce input groups=5  
Reduce shuffle bytes=260  
Reduce input records=24  
Reduce output records=5  
Spilled Records=48  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=120  
CPU time spent (ms)=1230  
Physical memory (bytes) snapshot=344330240  
Virtual memory (bytes) snapshot=3007062016  
Total committed heap usage (bytes)=226365440

#### 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=136

File Output Format Counters  
Bytes Written=34

*[cloudera@quickstart 33316]\$ hadoop fs -ls MRDir4*

Found 2 items

-rw-r--r--	1	cloudera	cloudera	0	2025-01-21 22:24	MRDir4/_SUCCESS
-rw-r--r--	1	cloudera	cloudera	34	2025-01-21 22:24	MRDir4/part-r-000000

*[cloudera@quickstart 33316]\$ hadoop fs -cat MRDir4/part-r-000000*

**BUS 7**

**CAR 7**

**RAJ 3**

**SOHAM 3**

**TRAIN 4**

[cloudera@quickstart 33316]\$