

P Sai Deekshith

BDA LAB CIE-2

Program-8 MapReduce (8a)

Program :

```
package countGender;

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());
```

P Sai Deekshith

BDA LAB CIE-2

Program-8 MapReduce (8a)

```
while (itr.hasMoreTokens()) {  
    word.set(itr.nextToken());  
    context.write(word, one);  
}  
}  
}
```

```
public static class IntSumReducer  
    extends Reducer<Text,IntWritable,Text,IntWritable> {  
    private IntWritable result = new IntWritable();  
  
    public void reduce(Text key, Iterable<IntWritable> values,  
        Context context  
        ) throws IOException, InterruptedException {  
        int sum = 0;  
  
        for (IntWritable val : values) {  
  
            sum += val.get();  
  
        }  
        if(key.toString() == "M" || key.toString() == "F") {  
            result.set(sum);  
  
            context.write(key, result);  
        }  
    }  
}
```

P Sai Deekshith

BDA LAB CIE-2

Program-8 MapReduce (8a)

```
}
```

```
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);  
}
```

```
}
```

P Sai Deekshith

BDA LAB CIE-2

Program-8 MapReduce (8a)

```
Activities Terminal Jun 1
hduser@psd

Map-Reduce Framework
  Map input records=5
  Map output records=10
  Map output bytes=77
  Map output materialized bytes=79
  Input split bytes=109
  Combine input records=10
  Combine output records=7
  Reduce input groups=7
  Reduce shuffle bytes=79
  Reduce input records=7
  Reduce output records=7
  Spilled Records=14
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=915
  CPU time spent (ms)=0
  Physical memory (bytes) snapshot=0
  Virtual memory (bytes) snapshot=0
  Total committed heap usage (bytes)=701497344

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=37
File Output Format Counters
  Bytes Written=45
hduser@psd-VirtualBox:~$ hdfs dfs -cat /labOutput/part-r-00000
21/06/14 16:12:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform..
F      2
Geetha 1
John   1
Kevin  1
M      3
Sita   1
Ton    1
hduser@psd-VirtualBox:~$ hadoop jar /home/psd/Desktop/CG2.jar countGender.WordCount /labInput/labInput
21/06/14 16:24:29 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform..
21/06/14 16:24:31 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.s
21/06/14 16:24:31 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId
21/06/14 16:24:31 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Impl
21/06/14 16:24:32 INFO input.FileInputFormat: Total input paths to process : 1
21/06/14 16:24:32 INFO mapreduce.JobSubmitter: number of splits:1
21/06/14 16:24:32 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local8686042_0001
21/06/14 16:24:33 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
21/06/14 16:24:33 INFO mapreduce.Job: Running job: job_local8686042_0001
21/06/14 16:24:33 INFO mapred.LocalJobRunner: OutputCommitter set in config null
```