

## Softcopy – Practical-1.4 (Hadoop)

### # Code

#### CharacterCountMapper.java

```
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

public class CharacterCountMapper extends Mapper<Object, Text, Text,
IntWritable> {

    private final static IntWritable one = new IntWritable(1);
    private Text character = new Text();

    @Override
    protected void map(Object key, Text value, Context context) throws
IOException, InterruptedException {
        String line = value.toString().toLowerCase();
        for (char c : line.toCharArray()) {
            if (Character.isAlphabetic(c)) {
                character.set(String.valueOf(c));
                context.write(character, one);
            }
        }
    }
}
```

#### CharacterCountReducer.java

```
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;
```

```

public class CharacterCountReducer extends Reducer<Text, IntWritable,
Text, IntWritable> {
    private IntWritable result = new IntWritable();

    @Override
    protected void reduce(Text key, Iterable<IntWritable> values, Context
context) throws IOException, InterruptedException {
        int count = 0;
        for (IntWritable val : values) {
            count += val.get();
        }
        result.set(count);
        context.write(key, result);
    }
}

```

### **CharacterCountDriver.java**

```

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 CharacterCountDriver {
    public static void main(String[] args) throws Exception {
        if (args.length != 2) {
            System.err.println("Usage: CharacterCountDriver <input path>
<output path>");
            System.exit(-1);
        }
    }
}

```

```

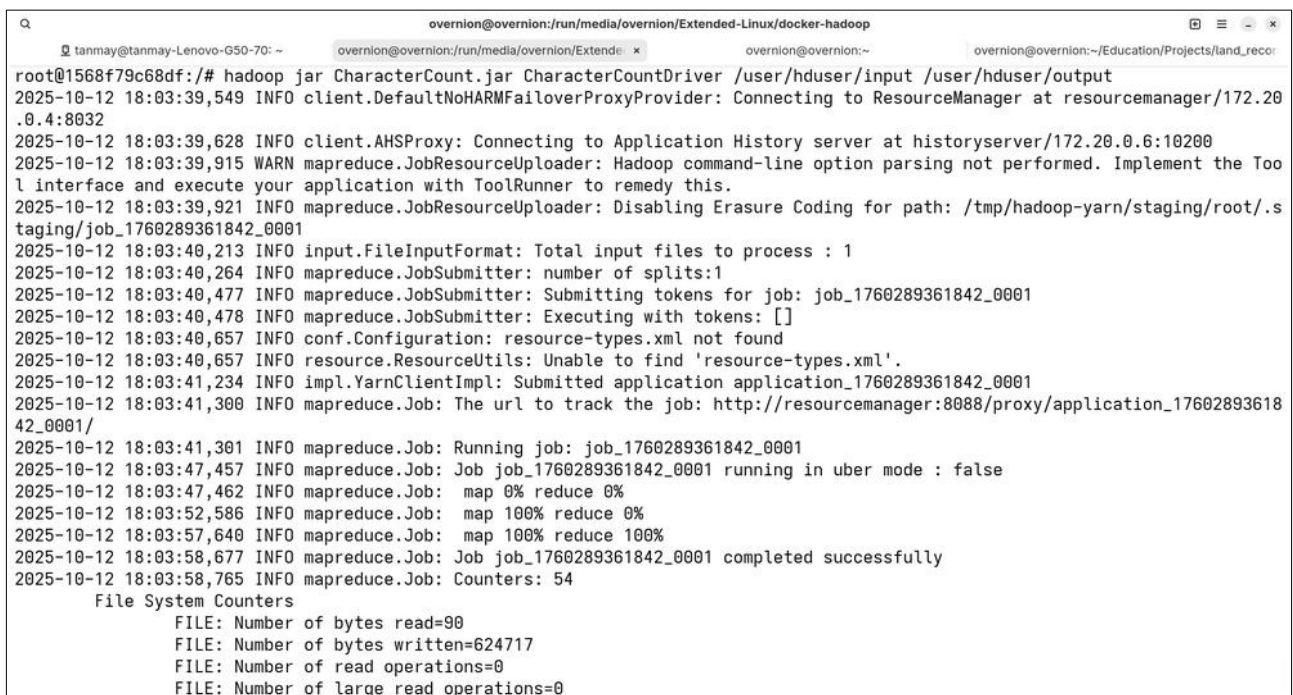
Configuration conf = new Configuration();
Job job = Job.getInstance(conf, "Character Count");
job.setJarByClass(CharacterCountDriver.class);
job.setMapperClass(CharacterCountMapper.class);
job.setCombinerClass(CharacterCountReducer.class);
job.setReducerClass(CharacterCountReducer.class);
job.setOutputKeyClass(Text.class);
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);
}
}

```

## # Output



```

overnion@overnion:/run/media/overnion/Extended-Linux/docker-hadoop
tanmay@tanmay-Lenovo-G50-70: ~
overnion@overnion:/run/media/overnion/Extended-Linux/docker-hadoop
overnion@overnion:~/Education/Projects/land_reco

root@1568f79c68df:/# hadoop jar CharacterCount.jar CharacterCountDriver /user/hduser/input /user/hduser/output
2025-10-12 18:03:39,549 INFO client.DefaultNoHARMAFailoverProxyProvider: Connecting to ResourceManager at resourcemanager/172.20.0.4:8032
2025-10-12 18:03:39,628 INFO client.AHSPProxy: Connecting to Application History server at historyserver/172.20.0.6:10200
2025-10-12 18:03:39,915 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2025-10-12 18:03:39,921 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/root/.staging/job_1760289361842_0001
2025-10-12 18:03:40,213 INFO input.FileInputFormat: Total input files to process : 1
2025-10-12 18:03:40,264 INFO mapreduce.JobSubmitter: number of splits:1
2025-10-12 18:03:40,477 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1760289361842_0001
2025-10-12 18:03:40,478 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-10-12 18:03:40,657 INFO conf.Configuration: resource-types.xml not found
2025-10-12 18:03:40,657 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-10-12 18:03:41,234 INFO impl.YarnClientImpl: Submitted application application_1760289361842_0001
2025-10-12 18:03:41,300 INFO mapreduce.Job: The url to track the job: http://resourcemanager:8088/proxy/application_1760289361842_0001/
2025-10-12 18:03:41,301 INFO mapreduce.Job: Running job: job_1760289361842_0001
2025-10-12 18:03:47,457 INFO mapreduce.Job: Job job_1760289361842_0001 running in uber mode : false
2025-10-12 18:03:47,462 INFO mapreduce.Job: map 0% reduce 0%
2025-10-12 18:03:52,586 INFO mapreduce.Job: map 100% reduce 0%
2025-10-12 18:03:57,640 INFO mapreduce.Job: map 100% reduce 100%
2025-10-12 18:03:58,677 INFO mapreduce.Job: Job job_1760289361842_0001 completed successfully
2025-10-12 18:03:58,765 INFO mapreduce.Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=90
        FILE: Number of bytes written=624717
        FILE: Number of read operations=0
        FILE: Number of large read operations=0

```

```

overnion@overnion:/run/media/overnion/Extended-Linux/docker-hadoop
tanmay@tanmay-Lenovo-G50-70: ~
overnion@overnion:/run/media/overnion/Extended-Linux/docker-hadoop
overnion@overnion:~/Education/Projects/land_reco:
2025-10-12 18:03:58,677 INFO mapreduce.Job: map 100% reduce 100%
2025-10-12 18:03:58,677 INFO mapreduce.Job: Job job_1760289361842_0001 completed successfully
2025-10-12 18:03:58,765 INFO mapreduce.Job: Counters: 54
File System Counters
  FILE: Number of bytes read=90
  FILE: Number of bytes written=624717
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=197
  HDFS: Number of bytes written=72
  HDFS: Number of read operations=8
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=1
  Launched reduce tasks=1
  Rack-local map tasks=1
  Total time spent by all maps in occupied slots (ms)=8656
  Total time spent by all reduces in occupied slots (ms)=16136
  Total time spent by all map tasks (ms)=2164
  Total time spent by all reduce tasks (ms)=2017
  Total vcore-milliseconds taken by all map tasks=2164
  Total vcore-milliseconds taken by all reduce tasks=2017
  Total megabyte-milliseconds taken by all map tasks=8863744
  Total megabyte-milliseconds taken by all reduce tasks=16523264
Map-Reduce Framework
  Map input records=4
  Map output records=65
  Map output bytes=390
  Map output materialized bytes=82
  Input split bytes=113
  Combine input records=65
  Combine output records=18
  Reduce input groups=18
  Reduce shuffle bytes=82
  Reduce input records=18
  Reduce output records=18
  Spilled Records=36
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=15
  CPU time spent (ms)=1200
  Physical memory (bytes) snapshot=433025024
  Virtual memory (bytes) snapshot=13759528960
  Total committed heap usage (bytes)=312475648
  Peak Map Physical memory (bytes)=262602752
  Peak Map Virtual memory (bytes)=5206544384
  Peak Reduce Physical memory (bytes)=170422272
  Peak Reduce Virtual memory (bytes)=8552984576
Shuffle Errors
  BAD_ID=0
  CONNECTION=0

```

```

overnion@overnion:/run/media/overnion/Extended-Linux/docker-hadoop
tanmay@tanmay-Lenovo-G50-70: ~
overnion@overnion:/run/media/overnion/Extended-Linux/docker-hadoop
overnion@overnion:~/Education/Projects/land_reco:
Total megabyte-milliseconds taken by all reduce tasks=16523264
Map-Reduce Framework
  Map input records=4
  Map output records=65
  Map output bytes=390
  Map output materialized bytes=82
  Input split bytes=113
  Combine input records=65
  Combine output records=18
  Reduce input groups=18
  Reduce shuffle bytes=82
  Reduce input records=18
  Reduce output records=18
  Spilled Records=36
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=15
  CPU time spent (ms)=1200
  Physical memory (bytes) snapshot=433025024
  Virtual memory (bytes) snapshot=13759528960
  Total committed heap usage (bytes)=312475648
  Peak Map Physical memory (bytes)=262602752
  Peak Map Virtual memory (bytes)=5206544384
  Peak Reduce Physical memory (bytes)=170422272
  Peak Reduce Virtual memory (bytes)=8552984576
Shuffle Errors
  BAD_ID=0
  CONNECTION=0

```

```

root@1568f79c68df:/# hdfs dfs -cat /user/hduser/output/part-r-00000
a      6
b      2
c      5
d      1
e      7
f      1
h      6
i      4
l      4
m      1
n      3
o      4
p      2
r      4
s      3
t      8
u      3
w      1

```