

MADHAV VEDPATHAK

Process.java

```
package com.mapreduce.lf;

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 Process{

    public static class IPMapper
        extends Mapper<Object, Text, Text, IntWritable>{

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

        public void map(Object key, Text value, Context context
```

```

        ) throws IOException, InterruptedException {
// Assuming the IP address is the first token in each line
StringTokenizer itr = new StringTokenizer(value.toString());
if (itr.hasMoreTokens()) {
    ip.set(itr.nextToken());
    context.write(ip, 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();
    }
    result.set(sum);
    context.write(key, result);
}
}

```

```

public static void main(String[] args) throws Exception {

    Configuration conf = new Configuration();

    Job job = Job.getInstance(conf, "IP address count");

    job.setJarByClass(Process.class);

    job.setMapperClass(IPMapper.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]));

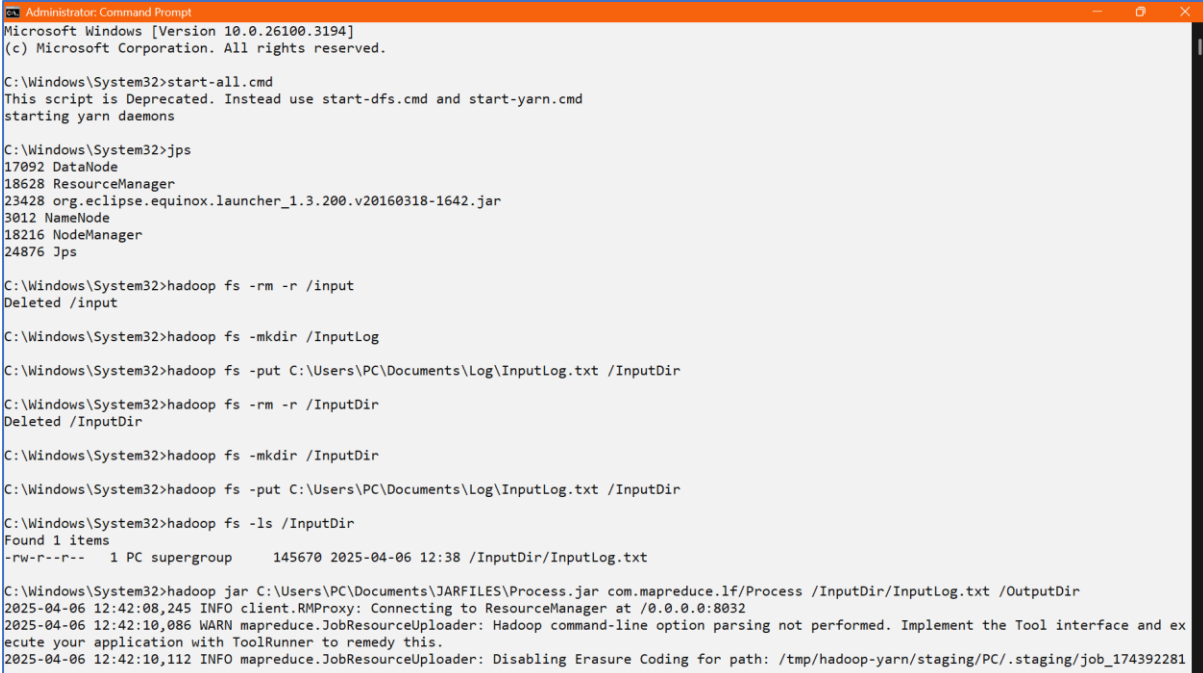
    System.exit(job.waitForCompletion(true) ? 0 : 1);

}

}

```

OUTPUTS:



```

Administrator: Command Prompt
Microsoft Windows [Version 10.0.26100.3194]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\System32>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

C:\Windows\System32>jps
17092 DataNode
18628 ResourceManager
23428 org.eclipse.equinox.launcher_1.3.200.v20160318-1642.jar
3012 NameNode
18216 NodeManager
24876 Jps

C:\Windows\System32>hadoop fs -rm -r /input
Deleted /input

C:\Windows\System32>hadoop fs -mkdir /InputLog

C:\Windows\System32>hadoop fs -put C:\Users\PC\Documents\Log\InputLog.txt /InputDir

C:\Windows\System32>hadoop fs -rm -r /InputDir
Deleted /InputDir

C:\Windows\System32>hadoop fs -mkdir /InputDir

C:\Windows\System32>hadoop fs -put C:\Users\PC\Documents\Log\InputLog.txt /InputDir

C:\Windows\System32>hadoop fs -ls /InputDir
Found 1 items
-rw-r--r-- 1 PC supergroup 145670 2025-04-06 12:38 /InputDir/InputLog.txt

C:\Windows\System32>hadoop jar C:\Users\PC\Documents\JARFILES\Process.jar com.mapreduce.lf/Process /InputDir/InputLog.txt /OutputDir
2025-04-06 12:42:08,245 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
2025-04-06 12:42:10,086 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and ex
ecute your application with ToolRunner to remedy this.
2025-04-06 12:42:10,112 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/PC/.staging/job_174392281

```

```
Administrator: Command Prompt
Found 1 items
-rw-r--r-- 1 PC supergroup 145670 2025-04-06 12:38 /InputDir/InputLog.txt

C:\Windows\System32>hadoop jar C:\Users\PC\Documents\JARFILES\Process.jar com.mapreduce.lf/Process /InputDir/InputLog.txt /OutputDir
2025-04-06 12:42:08,245 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2025-04-06 12:42:10,086 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and ex
ecute your application with ToolRunner to remedy this.
2025-04-06 12:42:10,112 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/PC/.staging/job_174392281
6858_0001
2025-04-06 12:42:10,675 INFO input.FileInputFormat: Total input files to process : 1
2025-04-06 12:42:10,851 INFO mapreduce.JobSubmitter: number of splits:1
2025-04-06 12:42:11,119 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1743922816858_0001
2025-04-06 12:42:11,124 INFO mapreduce.JobSubmitter: Executing with tokens: []
2025-04-06 12:42:11,569 INFO conf.Configuration: resource-types.xml not found
2025-04-06 12:42:11,571 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2025-04-06 12:42:12,407 INFO impl.YarnClientImpl: Submitted application application_1743922816858_0001
2025-04-06 12:42:12,517 INFO mapreduce.Job: The url to track the job: http://Vedpathak:8088/proxy/application_1743922816858_0001/
2025-04-06 12:42:12,521 INFO mapreduce.Job: Running job: job_1743922816858_0001
2025-04-06 12:42:31,946 INFO mapreduce.Job: Job job_1743922816858_0001 running in uber mode : false
2025-04-06 12:42:31,950 INFO mapreduce.Job: map 0% reduce 0%
2025-04-06 12:42:43,358 INFO mapreduce.Job: map 100% reduce 0%
2025-04-06 12:42:53,523 INFO mapreduce.Job: map 100% reduce 100%
2025-04-06 12:42:54,556 INFO mapreduce.Job: Job job_1743922816858_0001 completed successfully
2025-04-06 12:42:54,885 INFO mapreduce.Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=4479
        FILE: Number of bytes written=486695
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=145778
        HDFS: Number of bytes written=3611
        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
```

```
Administrator: Command Prompt

C:\Windows\System32>hadoop dfs -cat /OutputDir/*
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.
10.1.1.236 7
10.1.181.142 14
10.1.232.31 5
10.10.55.142 14
10.102.101.66 1
10.103.184.104 1
10.103.190.81 53
10.103.63.29 1
10.104.73.51 1
10.105.160.183 1
10.108.91.151 1
10.109.21.76 1
10.11.131.40 1
10.111.71.20 8
10.112.227.184 6
10.114.74.30 1
10.115.118.78 1
10.117.224.230 1
10.117.76.22 12
10.118.19.97 1
10.118.250.30 7
10.119.117.132 23
10.119.33.245 1
10.119.74.120 1
10.12.113.198 2
10.12.219.30 1
10.120.165.113 1
10.120.207.127 4
10.123.124.47 1
10.123.35.235 1
10.124.148.99 1
10.124.155.234 1
10.126.161.13 7
10.127.162.239 1
```

```
Administrator: Command Prompt

10.82.30.199 63
10.82.64.235 1
10.84.236.242 1
10.87.209.46 1
10.87.88.214 1
10.88.204.177 1
10.89.178.62 1
10.89.244.42 1
10.94.196.42 1
10.95.136.211 4
10.95.232.88 1
10.98.156.141 1
10.99.228.224 1

C:\Windows\System32>hdfs dfs -get /OutputDir/part-r-00000 C:\Users\PC\Documents\OutputDir\OutputLog.txt

C:\Windows\System32>
```

