

Sushma Hegde
1NT19IS169
C2 Batch

```
package sushma169;
import java.io.IOException;
import java.util.*;

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.mapred.FileInputFormat;
import org.apache.hadoop.mapred.FileOutputFormat;
import org.apache.hadoop.mapred.JobClient;
import org.apache.hadoop.mapred.JobConf;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reducer;
import org.apache.hadoop.mapred.Reporter;
import org.apache.hadoop.mapred.TextOutputFormat;
import org.apache.hadoop.mapred.TextInputFormat;
public class WordCount {
    public static class Map extends MapReduceBase implements
        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, OutputCollector<Text, IntWritable>
            output, Reporter reporter) throws IOException {
            String line = value.toString();
            StringTokenizer tokenizer = new StringTokenizer(line);
            while (tokenizer.hasMoreTokens()) {
                word.set(tokenizer.nextToken());
                output.collect(word, one);
            }
        }
    }

    public static class Reduce extends MapReduceBase implements
        Reducer<Text, IntWritable, Text, IntWritable> {

        public void reduce(Text key, Iterator<IntWritable> values,
            OutputCollector<Text, IntWritable> output, Reporter reporter)
```

```

throws IOException {
    int sum = 0;

    while (values.hasNext()) {
        sum += values.next().get();
    }
    output.collect(key, new IntWritable(sum));
}

}

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

    JobConf conf = new JobConf(WordCount.class);
    conf.setJobName("wordcount");
    conf.setOutputKeyClass(Text.class);
    conf.setOutputValueClass(IntWritable.class);
    conf.setMapperClass(Map.class);
    conf.setCombinerClass(Reduce.class);
    conf.setReducerClass(Reduce.class);
    conf.setInputFormat(TextInputFormat.class);
    conf.setOutputFormat(TextOutputFormat.class);
    FileInputFormat.setInputPaths(conf, new Path(args[0]));
    FileOutputFormat.setOutputPath(conf, new Path(args[1]));
    JobClient.runJob(conf);
}
}

```

```
Activities Terminal Jun 17 10:53
hdoop@admin1-HP-280-G4-MT-Business-PC: ~/hadoop-3.2.1/sbin

hdoop@admin1-HP-280-G4-MT-Business-PC:~$ jps
4617 Jps
hdoop@admin1-HP-280-G4-MT-Business-PC:~$ cd $HADOOP_HOME/sbin
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hdoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [admin1-HP-280-G4-MT-Business-PC]
Starting resource manager
Starting node managers
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ jps
5008 DataNode
5219 SecondaryNameNode
5566 NodeManager
4830 NameNode
5919 Jps
5407 ResourceManager
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -mkdir -p /input169
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -appendToFile - /input169/sush.txt
he who must not be named2022-06-17 10:39:55,388 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTru
sted = false
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hadoop jar ^C
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hadoop jar /home/hdoop/Desktop/sushma.jar /input169 /output169
2022-06-17 10:45:12,503 INFO client.RMPProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-17 10:45:12,637 INFO client.RMPProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-17 10:45:12,767 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute
your application with ToolRunner to remedy this.
2022-06-17 10:45:12,790 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_16554423186
35_0001
2022-06-17 10:45:12,879 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-17 10:45:12,980 INFO mapred.FileInputFormat: Total input files to process : 1
2022-06-17 10:45:13,023 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-17 10:45:13,082 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-17 10:45:13,497 INFO mapreduce.JobSubmitter: number of splits:2
2022-06-17 10:45:13,606 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-17 10:45:14,023 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1655442318635_0001
```

```
Activities Terminal Jun 17 10:46
hdoop@admin1-HP-280-G4-MT-Business-PC: ~/hadoop-3.2.1/sbin

Total megabyte-milliseconds taken by all reduce tasks=1591296
Map-Reduce Framework
  Map input records=1
  Map output records=6
  Map output bytes=49
  Map output materialized bytes=73
  Input split bytes=182
  Combine input records=6
  Combine output records=6
  Reduce input groups=6
  Reduce shuffle bytes=73
  Reduce input records=6
  Reduce output records=6
  Spilled Records=12
  Shuffled Maps =2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=152
  CPU time spent (ms)=1100
  Physical memory (bytes) snapshot=776814592
  Virtual memory (bytes) snapshot=7616004096
  Total committed heap usage (bytes)=8288999328
  Peak Map Physical memory (bytes)=294432768
  Peak Map Virtual memory (bytes)=2535473152
  Peak Reduce Physical memory (bytes)=192225280
  Peak Reduce Virtual memory (bytes)=2545602560
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=36
File Output Format Counters
  Bytes Written=37
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$
```

```
Activities Terminal Jun 17 10:47 hdoop@admin1-HP-280-G4-MT-Business-PC: ~/hadoop-3.2.1/sbin

Combine output records=6
Reduce input groups=6
Reduce shuffle bytes=73
Reduce input records=6
Reduce output records=6
Spilled Records=12
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=152
CPU time spent (ms)=1100
Physical memory (bytes) snapshot=776814592
Virtual memory (bytes) snapshot=7616004096
Total committed heap usage (bytes)=828899328
Peak Map Physical memory (bytes)=294432768
Peak Map Virtual memory (bytes)=2535473152
Peak Reduce Physical memory (bytes)=192225280
Peak Reduce Virtual memory (bytes)=2545602560

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=36
File Output Format Counters
Bytes Written=37
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -cat /output169/part*
2022-06-17 10:46:49,825 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
be 1
he 1
must 1
named 1
not 1
who 1
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$
```

```
Activities Terminal Jun 17 10:50 hdoop@admin1-HP-280-G4-MT-Business-PC: ~/hadoop-3.2.1/sbin

Peak Map Virtual memory (bytes)=2535473152
Peak Reduce Physical memory (bytes)=192225280
Peak Reduce Virtual memory (bytes)=2545602560

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=36
File Output Format Counters
Bytes Written=37
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -cat /output169/part*
2022-06-17 10:46:49,825 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
be 1
he 1
must 1
named 1
not 1
who 1
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -appendToFile - /input169/hegde.txt
hi how are you hi how are you i am fine i am fine
2022-06-17 10:48:39,941 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hadoop jar /home/hdoop/Desktop/hegde.jar /input169 /output169
JAR does not exist or is not a normal file: /home/hdoop/Desktop/hegde.jar
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hadoop jar /home/hdoop/Desktop/sushna.jar /input169 /output169
2022-06-17 10:49:36,719 INFO client.RMPProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-17 10:49:36,838 INFO client.RMPProxy: Connecting to ResourceManager at /127.0.0.1:8032
Exception in thread "main" org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://127.0.0.1:9000/output169 already exists
    at org.apache.hadoop.mapred.FileOutputFormat.checkOutputSpecs(FileOutputFormat.java:131)
    at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:279)
    at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:143)
    at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1570)
    at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1567)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://127.0.0.1:9000/output169 already exists
```

```
Activities Terminal Jun 17 10:50 hdoop@admin1-HP-280-G4-MT-Business-PC: ~/hadoop-3.2.1/sbin

Failed Shuffles=0
Merged Map outputs=3
GC time elapsed (ms)=302
CPU time spent (ms)=1440
Physical memory (bytes) snapshot=1165549568
Virtual memory (bytes) snapshot=10149273600
Total committed heap usage (bytes)=1184890880
Peak Map Physical memory (bytes)=341110784
Peak Map Virtual memory (bytes)=2535059456
Peak Reduce Physical memory (bytes)=191193088
Peak Reduce Virtual memory (bytes)=2544398336

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=87
File Output Format Counters
  Bytes Written=76
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -cat /output200/part*
2022-06-17 10:50:28,118 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
am      2
are     2
be      1
fine    2
he      1
hl      2
how     2
l       2
must    1
named   1
not     1
who     1
you     2
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$
```