Big Data Analytics (2CS702)

Practical 4	
Rollno: 19BCE055	Name: TEJAS DOBARIYA
Division: A	Batch: E1

AIM : Design MapReduce algorithms to take a very large file of integers and produce as output: 1. Wordcount 2. Largest integer 3. Average of numbers

1.WORDCOUNT.JAVA

```java
package wordcount;

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

```java
        }
    }

    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, "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);
    }
}
```

```
C:\Users\tejas>hadoop jar D:\7_Sem\BDA\wc.jar wordcount.wordcount /bda/word.txt /bda/out1
2022-10-12 20:44:34,280 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2022-10-12 20:44:34,551 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2022-10-12 20:44:34,552 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2022-10-12 20:44:35,938 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface a
2022-10-12 20:44:36,118 INFO input.FileInputFormat: Total input files to process : 1
2022-10-12 20:44:36,256 INFO mapreduce.JobSubmitter: number of splits:1
2022-10-12 20:44:36,482 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local96752893_0001
2022-10-12 20:44:36,483 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-10-12 20:44:36,702 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2022-10-12 20:44:36,705 INFO mapreduce.Job: Running job: job_local96752893_0001
2022-10-12 20:44:36,707 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2022-10-12 20:44:36,722 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2022-10-12 20:44:36,722 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, i
2022-10-12 20:44:36,724 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2022-10-12 20:44:36,837 INFO mapred.LocalJobRunner: Waiting for map tasks
2022-10-12 20:44:36,838 INFO mapred.LocalJobRunner: Starting task: attempt_local96752893_0001_m_000000_0
2022-10-12 20:44:36,888 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2022-10-12 20:44:36,889 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, i
2022-10-12 20:44:36,905 INFO util.ProcfsBasedProcessTree: ProcfsBasedProcessTree currently is supported only on Linux.
2022-10-12 20:44:36,988 INFO mapred.Task:  Using ResourceCalculatorProcessTree : org.apache.hadoop.yarn.util.WindowsBasedProcessTree@45224c6
2022-10-12 20:44:37,004 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/bda/word.txt:0+39
2022-10-12 20:44:37,152 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2022-10-12 20:44:37,152 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2022-10-12 20:44:37,152 INFO mapred.MapTask: soft limit at 83886080
2022-10-12 20:44:37,153 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2022-10-12 20:44:37,153 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2022-10-12 20:44:37,162 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2022-10-12 20:44:37,201 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-10-12 20:44:37,312 INFO mapred.LocalJobRunner:
2022-10-12 20:44:37,315 INFO mapred.MapTask: Starting flush of map output
2022-10-12 20:44:37,316 INFO mapred.MapTask: Spilling map output
2022-10-12 20:44:37,316 INFO mapred.MapTask: bufstart = 0; bufend = 70; bufvoid = 104857600
2022-10-12 20:44:37,316 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214368(104857472); length = 29/6553600
2022-10-12 20:44:37,347 INFO mapred.MapTask: Finished spill 0
2022-10-12 20:44:37,377 INFO mapred.Task: Task:attempt_local96752893_0001_m_000000_0 is done. And is in the process of committing
2022-10-12 20:44:37,385 INFO mapred.LocalJobRunner: map
2022-10-12 20:44:37,386 INFO mapred.Task: Task 'attempt_local96752893_0001_m_000000_0' done.
2022-10-12 20:44:37,401 INFO mapred.Task: Final Counters for attempt_local96752893_0001_m_000000_0: Counters: 24
        File System Counters
```

```
2022-10-12 20:44:37,402 INFO mapred.LocalJobRunner: Finishing task: attempt_local96752893_0001_m_000000_0
2022-10-12 20:44:37,405 INFO mapred.LocalJobRunner: map task executor complete.
2022-10-12 20:44:37,410 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2022-10-12 20:44:37,420 INFO mapred.LocalJobRunner: Starting task: attempt_local96752893_0001_r_000000_0
2022-10-12 20:44:37,438 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2022-10-12 20:44:37,438 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2022-10-12 20:44:37,441 INFO util.ProcfsBasedProcessTree: ProcfsBasedProcessTree currently is supported only on Linux.
2022-10-12 20:44:37,544 INFO mapred.Task:  Using ResourceCalculatorProcessTree : org.apache.hadoop.yarn.util.WindowsBasedProcessTree@faacad0
2022-10-12 20:44:37,553 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@14ef5367
2022-10-12 20:44:37,557 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2022-10-12 20:44:37,593 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=334338464, maxSingleShuffleLimit=83584616, mergeThreshold=220663392, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2022-10-12 20:44:37,616 INFO reduce.EventFetcher: attempt_local96752893_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2022-10-12 20:44:37,675 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local96752893_0001_m_000000_0 decomp: 68 len: 72 to MEMORY
2022-10-12 20:44:37,686 INFO reduce.InMemoryMapOutput: Read 68 bytes from map-output for attempt_local96752893_0001_m_000000_0
2022-10-12 20:44:37,690 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 68, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->68
2022-10-12 20:44:37,696 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
2022-10-12 20:44:37,698 INFO mapred.LocalJobRunner: 1 / 1 copied.
2022-10-12 20:44:37,704 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
2022-10-12 20:44:37,721 INFO mapreduce.Job: Job job_local96752893_0001 running in uber mode : false
2022-10-12 20:44:37,723 INFO mapreduce.Job:  map 100% reduce 0%
2022-10-12 20:44:37,731 INFO mapred.Merger: Merging 1 sorted segments
2022-10-12 20:44:37,732 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 60 bytes
2022-10-12 20:44:37,738 INFO reduce.MergeManagerImpl: Merged 1 segments, 68 bytes to disk to satisfy reduce memory limit
2022-10-12 20:44:37,744 INFO reduce.MergeManagerImpl: Merging 1 files, 72 bytes from disk
2022-10-12 20:44:37,747 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2022-10-12 20:44:37,753 INFO mapred.Merger: Merging 1 sorted segments
2022-10-12 20:44:37,763 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 60 bytes
2022-10-12 20:44:37,769 INFO mapred.LocalJobRunner: 1 / 1 copied.
2022-10-12 20:44:37,908 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords
2022-10-12 20:44:38,025 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-10-12 20:44:38,138 INFO mapred.Task: Task:attempt_local96752893_0001_r_000000_0 is done. And is in the process of committing
2022-10-12 20:44:38,143 INFO mapred.LocalJobRunner: 1 / 1 copied.
2022-10-12 20:44:38,144 INFO mapred.Task: Task attempt_local96752893_0001_r_000000_0 is allowed to commit now
2022-10-12 20:44:38,210 INFO output.FileOutputCommitter: Saved output of task 'attempt_local96752893_0001_r_000000_0' to hdfs://localhost:9000/bda/out1
2022-10-12 20:44:38,212 INFO mapred.LocalJobRunner: reduce > reduce
2022-10-12 20:44:38,220 INFO mapred.Task: Task 'attempt_local96752893_0001_r_000000_0' done.
2022-10-12 20:44:38,223 INFO mapred.Task: Final Counters for attempt_local96752893_0001_r_000000_0: Counters: 30
        File System Counters
            FILE: Number of bytes read=4442
```

```
            CONNECTION=0
            IO_ERROR=0
            WRONG_LENGTH=0
            WRONG_MAP=0
            WRONG_REDUCE=0
    File Output Format Counters
            Bytes Written=42
2022-10-12 20:44:38,225 INFO mapred.LocalJobRunner: Finishing task: attempt_local96752893_0001_r_000000_0
2022-10-12 20:44:38,227 INFO mapred.LocalJobRunner: reduce task executor complete.
2022-10-12 20:44:38,733 INFO mapreduce.Job:  map 100% reduce 100%
2022-10-12 20:44:38,734 INFO mapreduce.Job: Job job_local96752893_0001 completed successfully
2022-10-12 20:44:38,767 INFO mapreduce.Job: Counters: 36
    File System Counters
            FILE: Number of bytes read=8708
            FILE: Number of bytes written=1045824
            FILE: Number of read operations=0
            FILE: Number of large read operations=0
            FILE: Number of write operations=0
            HDFS: Number of bytes read=78
            HDFS: Number of bytes written=42
            HDFS: Number of read operations=15
            HDFS: Number of large read operations=0
            HDFS: Number of write operations=4
            HDFS: Number of bytes read erasure-coded=0
    Map-Reduce Framework
            Map input records=3
            Map output records=8
            Map output bytes=70
            Map output materialized bytes=72
            Input split bytes=99
            Combine input records=8
            Combine output records=6
            Reduce input groups=6
            Reduce shuffle bytes=72
            Reduce input records=6
            Reduce output records=6
            Spilled Records=12
            Shuffled Maps =1
            Failed Shuffles=0
            Merged Map outputs=1
            GC time elapsed (ms)=0
            Total committed heap usage (bytes)=548405248
    Shuffle Errors
            BAD_ID=0
            CONNECTION=0
```

```
*part-r-00000 - Notepad
File  Edit  Format  View  Help
Hello      1
Here       1
Hey        3
Hi         1
Nirma      1
University 1
There      1
```

2. INTMAX.JAVA

```java
package intMax;

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 intMax {

  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());
```

```java
        int max=Integer.MIN_VALUE;
        int num;
        while (itr.hasMoreTokens()) {
           num=Integer.parseInt(itr.nextToken());
           if(num>max) {
                   max=num;
           }
        }
        context.write(word, new IntWritable(max));
      }
    }

    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 max=Integer.MIN_VALUE;
        for (IntWritable val : values) {
          if(val.get()>max) {
                max=val.get();
          }
        }
        result.set(max);
        context.write(key, result);
      }
    }

    public static void main(String[] args) throws Exception {
      Configuration conf = new Configuration();
      Job job = Job.getInstance(conf, "word count");
      job.setJarByClass(intMax.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);
    }
}
```

integer.txt - Notepad

File Edit Format View Help

```
1 2 11 20 56 0 89 10 12
```

```
Select Command Prompt
C:\Users\tejas>hadoop fs -put D:\7_Sem\BDA\integer.txt /bda
2022-10-12 20:49:58,065 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false

C:\Users\tejas>hadoop jar D:\7_Sem\BDA\intMax.jar intMax.intMax /bda/integer.txt /bda/integerMax
2022-10-12 20:59:56,100 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2022-10-12 20:59:56,538 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2022-10-12 20:59:56,539 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2022-10-12 20:59:58,403 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRun
2022-10-12 20:59:58,557 INFO input.FileInputFormat: Total input files to process : 1
2022-10-12 20:59:58,843 INFO mapreduce.JobSubmitter: number of splits:1
2022-10-12 20:59:59,272 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local892819443_0001
2022-10-12 20:59:59,273 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-10-12 20:59:59,509 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2022-10-12 20:59:59,511 INFO mapreduce.Job: Running job: job_local892819443_0001
2022-10-12 20:59:59,515 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2022-10-12 20:59:59,534 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2022-10-12 20:59:59,536 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2022-10-12 20:59:59,538 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2022-10-12 20:59:59,691 INFO mapred.LocalJobRunner: Waiting for map tasks
2022-10-12 20:59:59,692 INFO mapred.LocalJobRunner: Starting task: attempt_local892819443_0001_m_000000_0
2022-10-12 20:59:59,742 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2022-10-12 20:59:59,742 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2022-10-12 20:59:59,764 INFO util.ProcfsBasedProcessTree: ProcfsBasedProcessTree currently is supported only on Linux.
2022-10-12 20:59:59,881 INFO mapred.Task:  Using ResourceCalculatorProcessTree : org.apache.hadoop.yarn.util.WindowsBasedProcessTree@2266be15
2022-10-12 20:59:59,919 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/bda/integer.txt:0+23
2022-10-12 21:00:00,120 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2022-10-12 21:00:00,121 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2022-10-12 21:00:00,122 INFO mapred.MapTask: soft limit at 83886080
2022-10-12 21:00:00,122 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2022-10-12 21:00:00,122 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2022-10-12 21:00:00,136 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2022-10-12 21:00:00,218 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-10-12 21:00:00,449 INFO mapred.LocalJobRunner:
2022-10-12 21:00:00,454 INFO mapred.MapTask: Starting flush of map output
2022-10-12 21:00:00,455 INFO mapred.MapTask: Spilling map output
2022-10-12 21:00:00,455 INFO mapred.MapTask: bufstart = 0; bufend = 5; bufvoid = 104857600
2022-10-12 21:00:00,455 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214396(104857584); length = 1/6553600
2022-10-12 21:00:00,482 INFO mapred.MapTask: Finished spill 0
2022-10-12 21:00:00,512 INFO mapred.Task: Task:attempt_local892819443_0001_m_000000_0 is done. And is in the process of committing
2022-10-12 21:00:00,520 INFO mapred.LocalJobRunner: map
2022-10-12 21:00:00,521 INFO mapred.Task: Task 'attempt_local892819443_0001_m_000000_0' done.
2022-10-12 21:00:00,537 INFO mapreduce.Job: Job job_local892819443_0001 running in uber mode : false
2022-10-12 21:00:00,539 INFO mapreduce.Job:   map 100% reduce 0%
2022-10-12 21:00:00,540 INFO mapred.Task: Final Counters for attempt_local892819443_0001_m_000000_0: Counters: 24
        File System Counters
                FILE: Number of bytes read=4320
```

```
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Output Format Counters
                Bytes Written=4
2022-10-12 21:00:01,407 INFO mapred.LocalJobRunner: Finishing task: attempt_local892819443_0001_r_000000_0
2022-10-12 21:00:01,408 INFO mapred.LocalJobRunner: reduce task executor complete.
2022-10-12 21:00:01,557 INFO mapreduce.Job:  map 100% reduce 100%
2022-10-12 21:00:01,557 INFO mapreduce.Job: Job job_local892819443_0001 completed successfully
2022-10-12 21:00:01,581 INFO mapreduce.Job: Counters: 36
        File System Counters
                FILE: Number of bytes read=8698
                FILE: Number of bytes written=1050775
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=46
                HDFS: Number of bytes written=4
                HDFS: Number of read operations=15
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=4
                HDFS: Number of bytes read erasure-coded=0
        Map-Reduce Framework
                Map input records=1
                Map output records=1
                Map output bytes=5
                Map output materialized bytes=13
                Input split bytes=102
                Combine input records=1
                Combine output records=1
                Reduce input groups=1
                Reduce shuffle bytes=13
                Reduce input records=1
                Reduce output records=1
                Spilled Records=2
                Shuffled Maps =1
                Failed Shuffles=0
```

prac4_2 - Notepad

File Edit Format View Help

Max Integer: 89

3. AVGINT.JAVA

```java
package avgInt;

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.FloatWritable;
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 avgInt {

  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());
      int count=0;
      int sum=0;
      int num;
      while (itr.hasMoreTokens()) {
        num=Integer.parseInt(itr.nextToken());
        count++;
        sum=sum=num;
      }
      float avg=sum/count;
      context.write(word, new IntWritable((int)avg));
    }
  }

  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;
      int count=0;
      for (IntWritable val : values) {
        count++;
```

```java
        sum+=val.get();
      }
      float avg=sum/count;
      result.set((int)avg);
      context.write(key, result);
    }
  }

  public static void main(String[] args) throws Exception {
    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, "word count");
    job.setJarByClass(avgInt.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);
  }
}
```

```
              Bytes Written=4

C:\Users\tejas>hadoop jar D:\7_Sem\BDA\avgInt.jar avgInt.avgInt /bda/integer.txt /bda/avgInt
2022-10-12 21:11:55,836 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2022-10-12 21:11:56,034 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2022-10-12 21:11:56,035 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2022-10-12 21:11:57,793 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2022-10-12 21:11:58,134 INFO input.FileInputFormat: Total input files to process : 1
2022-10-12 21:11:58,313 INFO mapreduce.JobSubmitter: number of splits:1
2022-10-12 21:11:58,568 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local233370023_0001
2022-10-12 21:11:58,568 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-10-12 21:11:58,851 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2022-10-12 21:11:58,854 INFO mapreduce.Job: Running job: job_local233370023_0001
2022-10-12 21:11:58,858 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2022-10-12 21:11:58,876 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2022-10-12 21:11:58,876 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2022-10-12 21:11:58,879 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2022-10-12 21:11:59,105 INFO mapred.LocalJobRunner: Waiting for map tasks
2022-10-12 21:11:59,106 INFO mapred.LocalJobRunner: Starting task: attempt_local233370023_0001_m_000000_0
2022-10-12 21:11:59,172 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2022-10-12 21:11:59,173 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2022-10-12 21:11:59,201 INFO util.ProcfsBasedProcessTree: ProcfsBasedProcessTree currently is supported only on Linux.
2022-10-12 21:11:59,325 INFO mapred.Task:  Using ResourceCalculatorProcessTree : org.apache.hadoop.yarn.util.WindowsBasedProcessTree@4916e4fc
2022-10-12 21:11:59,347 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/bda/integer.txt:0+23
2022-10-12 21:11:59,630 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
2022-10-12 21:11:59,631 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2022-10-12 21:11:59,632 INFO mapred.MapTask: soft limit at 83886080
2022-10-12 21:11:59,633 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2022-10-12 21:11:59,633 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2022-10-12 21:11:59,667 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
2022-10-12 21:11:59,807 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-10-12 21:11:59,875 INFO mapreduce.Job: Job job_local233370023_0001 running in uber mode : false
2022-10-12 21:11:59,880 INFO mapreduce.Job:  map 0% reduce 0%
2022-10-12 21:12:00,168 INFO mapred.LocalJobRunner:
2022-10-12 21:12:00,178 INFO mapred.MapTask: Starting flush of map output
2022-10-12 21:12:00,179 INFO mapred.MapTask: Spilling map output
2022-10-12 21:12:00,179 INFO mapred.MapTask: bufstart = 0; bufend = 5; bufvoid = 104857600
2022-10-12 21:12:00,181 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214396(104857584); length = 1/6553600
2022-10-12 21:12:00,245 INFO mapred.MapTask: Finished spill 0
2022-10-12 21:12:00,285 INFO mapred.Task: Task:attempt_local233370023_0001_m_000000_0 is done. And is in the process of committing
2022-10-12 21:12:00,303 INFO mapred.LocalJobRunner: map
2022-10-12 21:12:00,303 INFO mapred.Task: Task 'attempt_local233370023_0001_m_000000_0' done.
2022-10-12 21:12:00,349 INFO mapred.Task: Final Counters for attempt_local233370023_0001_m_000000_0: Counters: 24
        File System Counters
                FILE: Number of bytes read=4388
                FILE: Number of bytes written=525441
```

```
                Total committed heap usage (bytes)=275251200
        File Input Format Counters
                Bytes Read=23
2022-10-12 21:12:00,350 INFO mapred.LocalJobRunner: Finishing task: attempt_local233370023_0001_m_000000_0
2022-10-12 21:12:00,367 INFO mapred.LocalJobRunner: map task executor complete.
2022-10-12 21:12:00,380 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2022-10-12 21:12:00,381 INFO mapred.LocalJobRunner: Starting task: attempt_local233370023_0001_r_000000_0
2022-10-12 21:12:00,409 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 2
2022-10-12 21:12:00,410 INFO output.FileOutputCommitter: FileOutputCommitter skip cleanup _temporary folders under output directory:false, ignore cleanup failures: false
2022-10-12 21:12:00,417 INFO util.ProcfsBasedProcessTree: ProcfsBasedProcessTree currently is supported only on Linux.
2022-10-12 21:12:00,554 INFO mapred.Task:  Using ResourceCalculatorProcessTree : org.apache.hadoop.yarn.util.WindowsBasedProcessTree@7d214a9c
2022-10-12 21:12:00,563 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@5cf84ed
2022-10-12 21:12:00,569 WARN impl.MetricsSystemImpl: JobTracker metrics system already initialized!
2022-10-12 21:12:00,622 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=334338464, maxSingleShuffleLimit=83584616, mergeThreshold=220663392, ioSortFactor=10, memToMemMergeOutputsThreshold=10
2022-10-12 21:12:00,632 INFO reduce.EventFetcher: attempt_local233370023_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2022-10-12 21:12:00,726 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt_local233370023_0001_m_000000_0 decomp: 9 len: 13 to MEMORY
2022-10-12 21:12:00,756 INFO reduce.InMemoryMapOutput: Read 9 bytes from map-output for attempt_local233370023_0001_m_000000_0
2022-10-12 21:12:00,761 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 9, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->9
2022-10-12 21:12:00,765 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
2022-10-12 21:12:00,768 INFO mapred.LocalJobRunner: 1 / 1 copied.
2022-10-12 21:12:00,773 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
2022-10-12 21:12:00,815 INFO mapred.Merger: Merging 1 sorted segments
2022-10-12 21:12:00,816 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 6 bytes
2022-10-12 21:12:00,827 INFO reduce.MergeManagerImpl: Merged 1 segments, 9 bytes to disk to satisfy reduce memory limit
2022-10-12 21:12:00,831 INFO reduce.MergeManagerImpl: Merging 1 files, 13 bytes from disk
2022-10-12 21:12:00,836 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2022-10-12 21:12:00,837 INFO mapred.Merger: Merging 1 sorted segments
2022-10-12 21:12:00,849 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 6 bytes
2022-10-12 21:12:00,852 INFO mapred.LocalJobRunner: 1 / 1 copied.
2022-10-12 21:12:00,893 INFO mapreduce.Job:  map 100% reduce 0%
2022-10-12 21:12:01,483 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords
2022-10-12 21:12:01,962 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-10-12 21:12:02,536 INFO mapred.Task: Task:attempt_local233370023_0001_r_000000_0 is done. And is in the process of committing
2022-10-12 21:12:02,544 INFO mapred.LocalJobRunner: 1 / 1 copied.
2022-10-12 21:12:02,544 INFO mapred.Task: Task attempt_local233370023_0001_r_000000_0 is allowed to commit now
2022-10-12 21:12:02,687 INFO output.FileOutputCommitter: Saved output of task 'attempt_local233370023_0001_r_000000_0' to hdfs://localhost:9000/bda/avgInt
2022-10-12 21:12:02,689 INFO mapred.LocalJobRunner: reduce > reduce
2022-10-12 21:12:02,692 INFO mapred.Task: Task 'attempt_local233370023_0001_r_000000_0' done.
2022-10-12 21:12:02,701 INFO mapred.Task: Final Counters for attempt_local233370023_0001_r_000000_0: Counters: 30
        File System Counters
                                        WRONG_REDUCE=0
        File Output Format Counters
                Bytes Written=3
2022-10-12 21:12:02,703 INFO mapred.LocalJobRunner: Finishing task: attempt_local233370023_0001_r_000000_0
2022-10-12 21:12:02,705 INFO mapred.LocalJobRunner: reduce task executor complete.
2022-10-12 21:12:02,915 INFO mapreduce.Job:  map 100% reduce 100%
2022-10-12 21:12:02,916 INFO mapreduce.Job: Job job_local233370023_0001 completed successfully
2022-10-12 21:12:02,950 INFO mapreduce.Job: Counters: 36
        File System Counters
                FILE: Number of bytes read=8834
                FILE: Number of bytes written=1050895
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=46
                HDFS: Number of bytes written=3
                HDFS: Number of read operations=15
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=4
                HDFS: Number of bytes read erasure-coded=0
        Map-Reduce Framework
                Map input records=1
                Map output records=1
                Map output bytes=5
                Map output materialized bytes=13
                Input split bytes=102
                Combine input records=1
                Combine output records=1
                Reduce input groups=1
                Reduce shuffle bytes=13
                Reduce input records=1
                Reduce output records=1
                Spilled Records=2
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=0
                Total committed heap usage (bytes)=550502400
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
```

prac4_3 - Notepad

File Edit Format View Help

Avg: 22.3