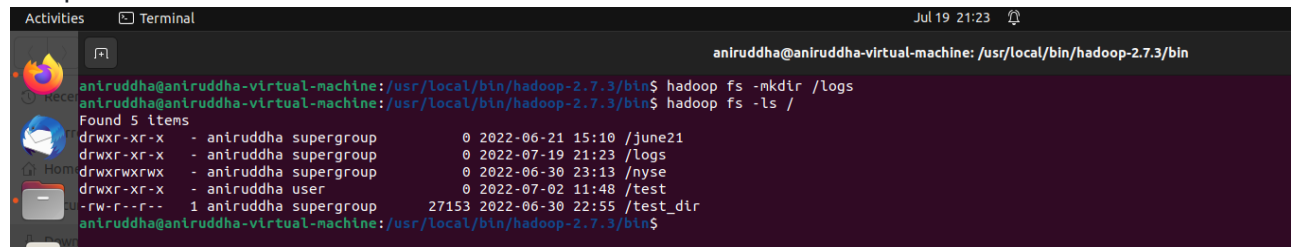


Aniruddha_Tambe_002101113_a04.md

PART 2 - Programming assignment

1. Creating /logs directory

Output:



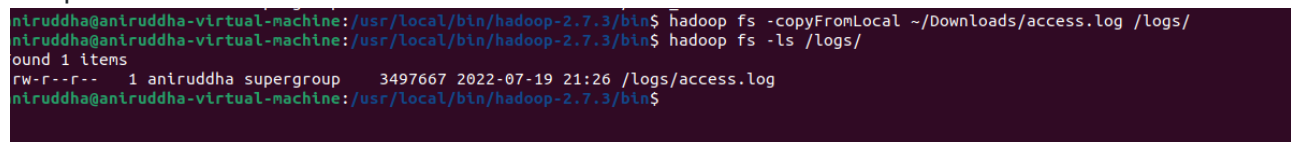
```

aniruddha@aniruddha-virtual-machine: /usr/local/bin/hadoop-2.7.3/bin
aniruddha@aniruddha-virtual-machine: /usr/local/bin/hadoop-2.7.3/bin$ hadoop fs -mkdir /logs
aniruddha@aniruddha-virtual-machine: /usr/local/bin/hadoop-2.7.3/bin$ hadoop fs -ls /
Found 5 items
drwxr-xr-x - aniruddha supergroup          0 2022-06-21 15:10 /june21
drwxr-xr-x - aniruddha supergroup          0 2022-07-19 21:23 /logs
drwxrwxrwx - aniruddha supergroup          0 2022-06-30 23:13 /nyse
drwxr-xr-x - aniruddha user                0 2022-07-02 11:48 /test
-rw-r--r-- 1 aniruddha supergroup    27153 2022-06-30 22:55 /test_dir
aniruddha@aniruddha-virtual-machine: /usr/local/bin/hadoop-2.7.3/bin$

```

2. Copy access.log file into HDFS

Output:



```

aniruddha@aniruddha-virtual-machine: /usr/local/bin/hadoop-2.7.3/bin$ hadoop fs -copyFromLocal ~/Downloads/access.log /logs/
aniruddha@aniruddha-virtual-machine: /usr/local/bin/hadoop-2.7.3/bin$ hadoop fs -ls /logs/
Found 1 items
-rw-r--r-- 1 aniruddha supergroup    3497667 2022-07-19 21:26 /logs/access.log
aniruddha@aniruddha-virtual-machine: /usr/local/bin/hadoop-2.7.3/bin$

```

3. Run jar file without combiner

```
package com.info7250.hadoop.accesslogs_mr;
```

```
import java.io.IOException;
```

```
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.input.TextInputFormat;
```

```
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
```

```
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
```

```
public class HitMain {
```

```
    public static void main(String args[]) throws IllegalArgumentException, IOEx
```

```
        Job job = Job.getInstance();
```

```
job.setJarByClass(HitMain.class);

// Specify various job-specific parameters
job.setMapperClass(HitCounter.class);
job.setReducerClass(HitReducer.class);

job.setInputFormatClass(TextInputFormat.class);
job.setOutputFormatClass(TextOutputFormat.class);

job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class);

FileInputFormat.addInputPath(job, new Path(args[0]));
FileOutputFormat.setOutputPath(job, new Path(args[1]));

job.setJobName("myjob");

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

}
```



Output:

```

22/07/23 18:27:40 INFO mapred.LocalJobRunner: Finishing task: attempt_local2006368503_0001_r_000
000_0
22/07/23 18:27:40 INFO mapred.LocalJobRunner: reduce task executor complete.
22/07/23 18:27:41 INFO mapreduce.Job: map 100% reduce 100%
22/07/23 18:27:41 INFO mapreduce.Job: Job job_local2006368503_0001 completed successfully
22/07/23 18:27:41 INFO mapreduce.Job: Counters: 35
    File System Counters
      FILE: Number of bytes read=1473736
      FILE: Number of bytes written=2769161
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=6995334
      HDFS: Number of bytes written=569063
      HDFS: Number of read operations=13
      HDFS: Number of large read operations=0
      HDFS: Number of write operations=4
    Map-Reduce Framework
      Map input records=35111
      Map output records=35111
      Map output bytes=639285
      Map output materialized bytes=709513
      Input split bytes=102
      Combine input records=0
      Combine output records=0
      Reduce input groups=1945
      Reduce shuffle bytes=709513
      Reduce input records=35111
      Reduce output records=35111
      Spilled Records=70222
      Shuffled Maps =1
      Failed Shuffles=0
      Merged Map outputs=1
      GC time elapsed (ms)=65
      Total committed heap usage (bytes)=550502400
    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=3497667
    File Output Format Counters
      Bytes Written=569063
aniruddha@aniruddha-virtual-machine: ~/Desktop/INF07250---Engineering-of-Big-Data-Systems/mongodb
$ 

```

4. Run jar with a combiner

```
package com.info7250.hadoop.accesslogs_mr;
```

```
import java.io.IOException;
```

```
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.input.TextInputFormat;
```

```
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
```

```
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
```

```
public class HitMain {  
  
    public static void main(String args[]) throws IllegalArgumentException, IOEx  
  
        Job job = Job.getInstance();  
        job.setJarByClass(HitMain.class);  
  
        // Specify various job-specific parameters  
        job.setMapperClass(HitCounter.class);  
        job.setCombinerClass(HitReducer.class);  
        job.setReducerClass(HitReducer.class);  
  
        job.setInputFormatClass(TextInputFormat.class);  
        job.setOutputFormatClass(TextOutputFormat.class);  
  
        job.setOutputKeyClass(Text.class);  
        job.setOutputValueClass(IntWritable.class);  
  
        FileInputFormat.addInputPath(job, new Path(args[0]));  
        FileOutputFormat.setOutputPath(job, new Path(args[1]));  
  
        job.setJobName("myjob");  
  
        System.exit(job.waitForCompletion(true) ? 0 :1);  
    }  
}
```



Output:

```

aniruddha@aniruddha-virtual-machine: ~/Desktop/INFO7250---Engineering-of-Big-...
22/07/23 18:34:25 INFO mapred.LocalJobRunner: Finishing task: attempt_local1307930030_0001_r_000
000_0
22/07/23 18:34:25 INFO mapred.LocalJobRunner: reduce task executor complete.
22/07/23 18:34:26 INFO mapreduce.Job: map 100% reduce 100%
22/07/23 18:34:26 INFO mapreduce.Job: Job job_local1307930030_0001 completed successfully
22/07/23 18:34:26 INFO mapreduce.Job: Counters: 35
  File System Counters
    FILE: Number of bytes read=1473774
    FILE: Number of bytes written=2770067
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=6995334
    HDFS: Number of bytes written=569063
    HDFS: Number of read operations=13
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=4
  Map-Reduce Framework
    Map input records=35111
    Map output records=35111
    Map output bytes=639285
    Map output materialized bytes=709513
    Input split bytes=102
    Combine input records=35111
    Combine output records=35111
    Reduce input groups=1945
    Reduce shuffle bytes=709513
    Reduce input records=35111
    Reduce output records=35111
    Spilled Records=70222
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=65
    Total committed heap usage (bytes)=550502400
  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=3497667
  File Output Format Counters
    Bytes Written=569063
aniruddha@aniruddha-virtual-machine: ~/Desktop/INFO7250---Engineering-of-Big-Data-Systems/mongodb
$

```

PART 3.1 – Copy NYSE dataset

Write a MapReduce to find the Max price of stock_price_high for each stock

Code:

```

package com.info7250.hadoop.stock;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.DoubleWritable;

```

```
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
import org.apache.hadoop.io.Text;

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

        Configuration configuration = new Configuration();

        Job job = Job.getInstance(configuration, "Ip Addr COUNTER");
        job.setJarByClass(StockDriver.class);

        job.setInputFormatClass(TextInputFormat.class);
        job.setOutputFormatClass(TextOutputFormat.class);

        job.setMapperClass(StockMapper.class);
        //job.setCombinerClass(IpReducer.class);
        job.setReducerClass(StockReducer.class);

        //output of the mapper
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(DoubleWritable.class);

        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));

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

Output for CombineFileInputFormat:

```

aniruddha@aniruddha-virtual-machine: ~/Desktop/INF07250---Engineering-of-Big-Data-Systems/mongodb
22/07/23 19:44:21 INFO mapred.LocalJobRunner: reduce > reduce
22/07/23 19:44:21 INFO mapred.Task: Task 'attempt_local2118019464_0001_r_000000_0' done.
22/07/23 19:44:21 INFO mapred.LocalJobRunner: Finishing task: attempt_local2118019464_0001_r_000000_0
22/07/23 19:44:21 INFO mapred.LocalJobRunner: reduce task executor complete.
22/07/23 19:44:21 INFO mapreduce.Job: map 100% reduce 100%
22/07/23 19:44:21 INFO mapreduce.Job: Job job_local2118019464_0001 completed successfully
22/07/23 19:44:21 INFO mapreduce.Job: Counters: 35
  File System Counters
    FILE: Number of bytes read=20564770
    FILE: Number of bytes written=31402427
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=81981984
    HDFS: Number of bytes written=1998
    HDFS: Number of read operations=13
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=4
  Map-Reduce Framework
    Map input records=735027
    Map output records=735026
    Map output bytes=8781587
    Map output materialized bytes=10251645
    Input split bytes=115
    Combine input records=0
    Combine output records=0
    Reduce input groups=203
    Reduce shuffle bytes=10251645
    Reduce input records=735026
    Reduce output records=203
    Spilled Records=1470052
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=43
    Total committed heap usage (bytes)=621805568
  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=40990992
  File Output Format Counters
    Bytes Written=1998
aniruddha@aniruddha-virtual-machine: ~/Desktop/INF07250---Engineering-of-Big-Data-Systems/mongodb$

```

PART 3.2 – Redo 3.1

Individual file elapsed time: 32 seconds

```

aniruddha@aniruddha-virtual-machine: ~/Desktop/INFO7250---Engineering-of-Big-Data-Systems/mongodb
22/07/23 19:44:21 INFO mapred.LocalJobRunner: reduce > reduce
22/07/23 19:44:21 INFO mapred.Task: Task 'attempt_local2118019464_0001_r_000000_0' done.
22/07/23 19:44:21 INFO mapred.LocalJobRunner: Finishing task: attempt_local2118019464_0001_r_000000_0
22/07/23 19:44:21 INFO mapred.LocalJobRunner: reduce task executor complete.
22/07/23 19:44:21 INFO mapreduce.Job: map 100% reduce 100%
22/07/23 19:44:21 INFO mapreduce.Job: Job job_local2118019464_0001 completed successfully
22/07/23 19:44:21 INFO mapreduce.Job: Counters: 35
  File System Counters
    FILE: Number of bytes read=20564770
    FILE: Number of bytes written=31402427
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=81981984
    HDFS: Number of bytes written=1998
    HDFS: Number of read operations=13
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=4
  Map-Reduce Framework
    Map input records=735027
    Map output records=735026
    Map output bytes=8781587
    Map output materialized bytes=10251645
    Input split bytes=115
    Combine input records=0
    Combine output records=0
    Reduce input groups=203
    Reduce shuffle bytes=10251645
    Reduce input records=735026
    Reduce output records=203
    Spilled Records=1470052
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=43
    Total committed heap usage (bytes)=621805568
  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=40990992
  File Output Format Counters
    Bytes Written=27922
aniruddha@aniruddha-virtual-machine: ~/Desktop/INFO7250---Engineering-of-Big-Data-Systems/mongodb$

```

Merged file elapsed time : 25 seconds

```

22/07/23 20:14:35 INFO mapred.LocalJobRunner: Finishing task: attempt_local1954245375_0001_r_000000_0
22/07/23 20:14:35 INFO mapred.LocalJobRunner: reduce task executor complete.
22/07/23 20:14:36 INFO mapreduce.Job: map 100% reduce 100%
22/07/23 20:14:36 INFO mapreduce.Job: Job job_local1954245375_0001 completed successfully
22/07/23 20:14:36 INFO mapreduce.Job: Counters: 35
  File System Counters
    FILE: Number of bytes read=256506035
    FILE: Number of bytes written=586947162
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=1834364840
    HDFS: Number of bytes written=27922
    HDFS: Number of read operations=46
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=7
  Map-Reduce Framework
    Map input records=9343831
    Map output records=9211031
    Map output bytes=109751285
    Map output materialized bytes=128173371
    Input split bytes=388
    Combine input records=0
    Combine output records=0
    Reduce input groups=2853
    Reduce shuffle bytes=128173371
    Reduce input records=9211031
    Reduce output records=2853
    Spilled Records=18422062
    Shuffled Maps =4
    Failed Shuffles=0
    Merged Map outputs=4
    GC time elapsed (ms)=515
    Total committed heap usage (bytes)=1939341312
  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=514516948
  File Output Format Counters
    Bytes Written=27922

```

PART 4 – Programming Assignment

CombineFileInputFormat, FixedLengthInputFormat, KeyValueTypeInputFormat, NLineInputFormat & TextInputFormat

Code for Combined My Format:

```
package com.info7250.hadoop.customInputFormat;

import java.io.IOException;

import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.InputSplit;
import org.apache.hadoop.mapreduce.RecordReader;
import org.apache.hadoop.mapreduce.TaskAttemptContext;
import org.apache.hadoop.mapreduce.lib.input.CombineFileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.CombineFileRecordReader;
import org.apache.hadoop.mapreduce.lib.input.CombineFileRecordReaderWrapper;
import org.apache.hadoop.mapreduce.lib.input.CombineFileSplit;
import org.apache.hadoop.mapreduce.lib.input.FileSplit;
import org.apache.hadoop.mapreduce.lib.input.LineRecordReader;

public class CombinedMyFormat extends CombineFileInputFormat<LongWritable,Text>{

    /*private static class CombineMyKeyMyValueReaderWrapper
    extends CombineFileRecordReaderWrapper< Text, Text > {
        protected CombineMyKeyMyValueReaderWrapper(
            CombineFileSplit split, TaskAttemptContext ctx, Integer idx
        ) throws IOException, InterruptedException {
            super( new CombinedMyFormat(), split, ctx, idx );
        }
    }

    @Override
    public RecordReader<Text, Text> createRecordReader(InputSplit split, TaskAtt
        throws IOException {
        // TODO Auto-generated method stub
        return new CombineFileRecordReader< Text, Text >(
            ( CombineFileSplit )split, ctx, CombineMyKeyMyValueReaderWrapper
        }*/

    @Override
    public RecordReader<LongWritable, Text>
        createRecordReader(InputSplit split, TaskAttemptContext context)
            throws IOException {

        CombineFileRecordReader<LongWritable, Text> reader =
            new CombineFileRecordReader<LongWritable, Text>(
                (CombineFileSplit) split, context, myCombineFileRecordReader
            return reader;
        }
    }
```

```
public static class myCombineFileRecordReader extends RecordReader<LongWritable,
    private LineRecordReader lineRecordReader = new LineRecordReader();

    public myCombineFileRecordReader(CombineFileSplit split,
        TaskAttemptContext context, Integer index) throws IOException {

        FileSplit fileSplit = new FileSplit(split.getPath(index),
            split.getOffset(index),
            split.getLength(index),
            split.getLocations());
        lineRecordReader.initialize(fileSplit, context);
    }

    @Override
    public void initialize(InputSplit inputSplit, TaskAttemptContext context)
        throws IOException, InterruptedException {
        //linerecordReader.initialize(inputSplit, context);
    }

    @Override
    public void close() throws IOException {
        lineRecordReader.close();
    }

    @Override
    public float getProgress() throws IOException {
        return lineRecordReader.getProgress();
    }

    @Override
    public LongWritable getCurrentKey() throws IOException,
        InterruptedException {
        return lineRecordReader.getCurrentKey();
    }

    @Override
    public Text getCurrentValue() throws IOException, InterruptedException {
        return lineRecordReader.getCurrentValue();
    }

    @Override
    public boolean nextKeyValue() throws IOException, InterruptedException {
        return lineRecordReader.nextKeyValue();
    }
}
```

}

Code for driver class:

```
package com.info7250.hadoop.customInputFormat;

import java.io.IOException;

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.input.FixedLengthInputFormat;
import org.apache.hadoop.mapreduce.lib.input.KeyValueTextInputFormat;
import org.apache.hadoop.mapreduce.lib.input.NLineInputFormat;
//import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;

public class HitMain {

    public static void main(String args[]) throws IllegalArgumentException, IOEx

    //***** FixedLengthInputFormat *****

        Configuration conf = new Configuration();
        conf.set("mapreduce.input.keyvaluelinerecordreader.key.value.separat
        //conf.setInt(FixedLengthInputFormat.FIXED_RECORD_LENGTH, 11);
        conf.setInt("fixedlengthinputformat.record.length",8);

        Job job = new Job(conf);
        //Job job = Job.getInstance();
        job.setJarByClass(HitMain.class);

        // Specify various job-specific parameters
        job.setMapperClass(HitCounter.class);
        job.setReducerClass(HitReducer.class);

        job.setInputFormatClass(FixedLengthInputFormat.class);
        //FixedLengthInputFormat.setRecordLength(conf, 100);

    //***** KeyValueTextInputFormat *****

        /*Configuration conf = new Configuration();
        conf.set("mapreduce.input.keyvaluelinerecordreader.key.value.separat

        Job job = new Job(conf);
```

```
//Job job = Job.getInstance();
job.setJarByClass(HitMain.class);

// Specify various job-specific parameters
job.setMapperClass(HitCounter.class);
job.setReducerClass(HitReducer.class);

//job.setInputFormatClass(KeyValueTextInputFormat.class);*/

//***** NLineInputFormat *****

//Configuration conf = new Configuration();
//conf.set("mapreduce.input.keyvaluelinerecordreader.key.value.separ

//Job job = new Job(conf);
/*Job job = Job.getInstance();
job.getConfiguration().setInt("mapreduce.input.lineinputformat.lines
job.setJarByClass(HitMain.class);

// Specify various job-specific parameters
job.setMapperClass(HitCounter.class);
job.setReducerClass(HitReducer.class);

job.setInputFormatClass(NLineInputFormat.class);*/

//***** CombineTextInputFormat *****
/*Configuration conf = new Configuration();
conf.set("mapred.textoutputformat.separator", ",");
conf.set("mapred.max.split.size", "16777216");

Job job = new Job(conf);
job.getConfiguration().setLong("mapreduce.input.fileinputformat.split.maxsiz

job.setJarByClass(HitMain.class);
job.setMapperClass(HitCounter.class);
job.setReducerClass(HitReducer.class);

job.setInputFormatClass(CombinedMyFormat.class);
CombinedMyFormat.setMaxInputSplitSize(job, 16777216);*/

// *****
job.setOutputFormatClass(TextOutputFormat.class);

job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class);

FileInputFormat.addInputPath(job, new Path(args[0]));
FileOutputFormat.setOutputPath(job, new Path(args[1]));
```

```
        job.setJobName("myjob");

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

}
```

Code for Mapper:

```
package com.info7250.hadoop.customInputFormat;

import java.io.IOException;
import java.nio.charset.StandardCharsets;

import org.apache.hadoop.io.BytesWritable;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

//public class HitCounter extends Mapper<Text,Text,Text,IntWritable>{ //- KeyValueTe
//public class HitCounter extends Mapper<LongWritable,Text,Text,IntWritable>{ //- NL
public class HitCounter extends Mapper<LongWritable,BytesWritable,Text,IntWritable>{

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

    //public void map(Text key, Text value, org.apache.hadoop.mapreduce.Mapper.C
    //public void map(BytesWritable key, Text value, org.apache.hadoop.mapreduce
    public void map(LongWritable key, BytesWritable value, org.apache.hadoop.map

        //String line=value.toString();
        String line = new String(value.get(),StandardCharsets.UTF_8)

        String[] tokens=line.split(" ");
        ipaddress.set(tokens[0]);

        context.write(ipaddress, one);
        //context gives output key and value
    }

}
```

Output for CombineFileInputFormat:

```

aniruddha@aniruddha-virtual-machine: ~/Desktop/INFO7250---Engineering-of-Big-...
22/07/23 17:48:50 INFO mapred.Task: Task 'attempt_local1323112930_0001_r_000000_0' done.
22/07/23 17:48:50 INFO mapred.LocalJobRunner: Finishing task: attempt_local1323112930_0001_r_000
000_0
22/07/23 17:48:50 INFO mapred.LocalJobRunner: reduce task executor complete.
22/07/23 17:48:50 INFO mapreduce.Job: Job job_local1323112930_0001 running in uber mode : false
22/07/23 17:48:50 INFO mapreduce.Job: map 100% reduce 100%
22/07/23 17:48:50 INFO mapreduce.Job: Job job_local1323112930_0001 completed successfully
22/07/23 17:48:50 INFO mapreduce.Job: Counters: 35
    File System Counters
        FILE: Number of bytes read=55024
        FILE: Number of bytes written=642718
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=192
        HDFS: Number of bytes written=92
        HDFS: Number of read operations=13
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=4
    Map-Reduce Framework
        Map input records=12
        Map output records=12
        Map output bytes=116
        Map output materialized bytes=146
        Input split bytes=100
        Combine input records=0
        Combine output records=0
        Reduce input groups=5
        Reduce shuffle bytes=146
        Reduce input records=12
        Reduce output records=12
        Spilled Records=24
        Shuffled Maps =1
        Failed Shuffles=0
        Merged Map outputs=1
        GC time elapsed (ms)=0
        Total committed heap usage (bytes)=557842432
    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=96
    File Output Format Counters
        Bytes Written=92

```

Output for FixedLengthInputFormat:

```
aniruddha@aniruddha-virtual-machine: ~/Desktop/INFO7250---Engineering-of-Big-...
22/07/23 18:27:40 INFO mapred.LocalJobRunner: reduce task executor complete.
22/07/23 18:27:41 INFO mapreduce.Job: map 100% reduce 100%
22/07/23 18:27:41 INFO mapreduce.Job: Job job_local2006368503_0001 completed successfully
22/07/23 18:27:41 INFO mapreduce.Job: Counters: 35
  File System Counters
    FILE: Number of bytes read=1473736
    FILE: Number of bytes written=2769161
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=6995334
    HDFS: Number of bytes written=569063
    HDFS: Number of read operations=13
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=4
  Map-Reduce Framework
    Map input records=35111
    Map output records=35111
    Map output bytes=639285
    Map output materialized bytes=709513
    Input split bytes=102
    Combine input records=0
    Combine output records=0
    Reduce input groups=1945
    Reduce shuffle bytes=709513
    Reduce input records=35111
    Reduce output records=35111
    Spilled Records=70222
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=65
    Total committed heap usage (bytes)=550502400
  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=3497667
  File Output Format Counters
    Bytes Written=569063
aniruddha@aniruddha-virtual-machine: ~/Desktop/INFO7250---Engineering-of-Big-Data-Systems/mongodb
```

Output for NLineInputFormat:

```

22/07/23 16:57:57 INFO mapred.LocalJobRunner: Finishing task: attempt_local2074151326_0001_r_000000_0
22/07/23 16:57:57 INFO mapred.LocalJobRunner: reduce task executor complete.
22/07/23 16:57:58 INFO mapreduce.Job: map 100% reduce 100%
22/07/23 16:57:58 INFO mapreduce.Job: Job job_local2074151326_0001 completed successfully
22/07/23 16:57:58 INFO mapreduce.Job: Counters: 35
  File System Counters
    FILE: Number of bytes read=1473788
    FILE: Number of bytes written=2770467
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=6995334
    HDFS: Number of bytes written=569063
    HDFS: Number of read operations=13
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=4
  Map-Reduce Framework
    Map input records=35111
    Map output records=35111
    Map output bytes=639285
    Map output materialized bytes=709513
    Input split bytes=129
    Combine input records=0
    Combine output records=0
    Reduce input groups=1945
    Reduce shuffle bytes=709513
    Reduce input records=35111
    Reduce output records=35111
    Spilled Records=70222
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=19
    Total committed heap usage (bytes)=562036736
  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=0
  File Output Format Counters
    Bytes Written=569063

```

Output for KeyValueTextInputFormat:

```

22/07/23 16:57:57 INFO mapred.LocalJobRunner: Finishing task: attempt_local2074151326_0001_r_000000_0
22/07/23 16:57:57 INFO mapred.LocalJobRunner: reduce task executor complete.
22/07/23 16:57:58 INFO mapreduce.Job: map 100% reduce 100%
22/07/23 16:57:58 INFO mapreduce.Job: Job job_local2074151326_0001 completed successfully
22/07/23 16:57:58 INFO mapreduce.Job: Counters: 35
  File System Counters
    FILE: Number of bytes read=1473788
    FILE: Number of bytes written=2770467
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=6995334
    HDFS: Number of bytes written=569063
    HDFS: Number of read operations=13
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=4
  Map-Reduce Framework
    Map input records=35111
    Map output records=35111
    Map output bytes=639285
    Map output materialized bytes=709513
    Input split bytes=129
    Combine input records=0
    Combine output records=0
    Reduce input groups=1945
    Reduce shuffle bytes=709513
    Reduce input records=35111
    Reduce output records=35111
    Spilled Records=70222
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=19
    Total committed heap usage (bytes)=562036736
  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=0
  File Output Format Counters
    Bytes Written=569063

```


Output for TextInputFormat:

```

22/07/23 18:34:25 INFO mapred.LocalJobRunner: Finishing task: attempt_local1307930030_0001_r_000
000_0
22/07/23 18:34:25 INFO mapred.LocalJobRunner: reduce task executor complete.
22/07/23 18:34:26 INFO mapreduce.Job: map 100% reduce 100%
22/07/23 18:34:26 INFO mapreduce.Job: Job job_local1307930030_0001 completed successfully
22/07/23 18:34:26 INFO mapreduce.Job: Counters: 35
    File System Counters
        FILE: Number of bytes read=1473774
        FILE: Number of bytes written=2770067
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=6995334
        HDFS: Number of bytes written=569063
        HDFS: Number of read operations=13
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=4
    Map-Reduce Framework
        Map input records=35111
        Map output records=35111
        Map output bytes=639285
        Map output materialized bytes=709513
        Input split bytes=102
        Combine input records=35111
        Combine output records=35111
        Reduce input groups=1945
        Reduce shuffle bytes=709513
        Reduce input records=35111
        Reduce output records=35111
        Spilled Records=70222
        Shuffled Maps =1
        Failed Shuffles=0
        Merged Map outputs=1
        GC time elapsed (ms)=65
        Total committed heap usage (bytes)=550502400
    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=3497667
    File Output Format Counters
        Bytes Written=569063
aniruddha@aniruddha-virtual-machine: ~/Desktop/INFO7250---Engineering-of-Big-Data-Systems/mongodb
$

```

Part 5

Code:

```

package com.info7250.hadoop.stocks;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.DoubleWritable;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;

```

```
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;

import org.apache.hadoop.io.Text;

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

        Configuration configuration = new Configuration();

        Job job = Job.getInstance(configuration, "Ip Addr COUNTER");
        job.setJarByClass(DateStockDriver.class);

        job.setInputFormatClass(TextInputFormat.class);
        job.setOutputFormatClass(TextOutputFormat.class);

        /*job.setMapperClass(DateStockMapper.class);
        //job.setCombinerClass(IpReducer.class);
        job.setReducerClass(DateStockReducer.class);

        //output of the mapper
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(Text.class);

        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));

        System.exit(job.waitForCompletion(true) ? 0 : 1);*/
        //Mapper and Reducer for Part 5
        job.setMapperClass(Part5Mapper.class);
        //job.setCombinerClass(Part5Reducer.class);
        job.setReducerClass(Part5Reducer.class);

        //Output Types for Part 5
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(NYSEWritable.class);

        job.setInputFormatClass(TextInputFormat.class);
        job.setOutputFormatClass(TextOutputFormat.class);

        FileInputFormat.setInputPaths(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));
        System.exit(job.waitForCompletion(true)? 0: 1);
    }
}
```

Output for TextInputFormat:

```

22/07/23 15:04:22 INFO mapred.LocalJobRunner: reduce > reduce
22/07/23 15:04:22 INFO mapred.Task: Task 'attempt_local1809411942_0001_r_000000_0' done.
22/07/23 15:04:22 INFO mapred.LocalJobRunner: Finishing task: attempt_local1809411942_0001_r_000000_0
22/07/23 15:04:22 INFO mapred.LocalJobRunner: reduce task executor complete.
22/07/23 15:04:22 INFO mapreduce.Job: map 100% reduce 100%
22/07/23 15:04:22 INFO mapreduce.Job: Job job_local1809411942_0001 completed successfully
22/07/23 15:04:22 INFO mapreduce.Job: Counters: 35
  File System Counters
    FILE: Number of bytes read=73470356
    FILE: Number of bytes written=110756361
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=81981724
    HDFS: Number of bytes written=8557
    HDFS: Number of read operations=13
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=4
  Map-Reduce Framework
    Map input records=735026
    Map output records=735026
    Map output bytes=35242523
    Map output materialized bytes=36712581
    Input split bytes=110
    Combine input records=0
    Combine output records=0
    Reduce input groups=203
    Reduce shuffle bytes=36712581
    Reduce input records=735026
    Reduce output records=203
    Spilled Records=1470052
    Shuffled Maps=1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=389
    Total committed heap usage (bytes)=625016832
  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=40990862
  File Output Format Counters
    Bytes Written=8557
sid@sid-virtual-machine: /usr/local/bin/hadoop-2.7.3/sbin$

```

Part 6

Output for TextInputFormat:

```

22/07/23 15:12:35 INFO mapred.Task: Task 'attempt_local1583716670_0001_r_000000_0' done.
22/07/23 15:12:35 INFO mapred.LocalJobRunner: Finishing task: attempt_local1583716670_0001_r_000000_0
22/07/23 15:12:35 INFO mapreduce.Job: map 100% reduce 100%
22/07/23 15:12:36 INFO mapreduce.Job: Job job_local1583716670_0001 completed successfully
22/07/23 15:12:36 INFO mapreduce.Job: Counters: 35
  File System Counters
    FILE: Number of bytes read=65536
    FILE: Number of bytes written=652922
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=81981724
    HDFS: Number of bytes written=8557
    HDFS: Number of read operations=13
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=4
  Map-Reduce Framework
    Map input records=735026
    Map output records=735026
    Map output bytes=35242523
    Map output materialized bytes=10150
    Input split bytes=110
    Combine input records=735026
    Combine output records=203
    Reduce input groups=203
    Reduce shuffle bytes=10150
    Reduce input records=203
    Reduce output records=203
    Spilled Records=406
    Shuffled Maps=1
    Failed Shuffles=0
    Merged Map outputs=1
    GC time elapsed (ms)=282
    Total committed heap usage (bytes)=634388480
  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=40990862
  File Output Format Counters
    Bytes Written=8557

```