Advance map reduce on Access.csv file

1] Find the Top 10 most visiting IP Address

Mapper

```java
package sayali.Assignment4Part6;

import java.io.IOException;
import java.util.List;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;
import com.google.common.base.Splitter;
import com.google.common.collect.Lists;

public class top10IpAddressMapper extends Mapper<Object, Text, Text, LongWritable>
    {
    @Override
    public void map(Object key, Text value, Context context)throws IOException,
InterruptedException {
                LongWritable one = new LongWritable(1);
                String line = value.toString();
                List<String> items = Lists.newArrayList(Splitter.on(',').split(line));
            String ipaddress = items.get(0);
            Text Key = new Text(ipaddress);
            context.write(Key,one);
        } }
```

Reducer:

```java
package sayali.Assignment4Part6;
import java.io.IOException;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

public class top10IpAddressReducer extends Reducer<Text, LongWritable, Text,
LongWritable> {

    public void reduce(Text key, Iterable<LongWritable> values, Context context) throws
IOException, InterruptedException
        {
```

```java
                int sum = 0;
                for (LongWritable v : values)
                {
                        sum += v.get();
    }
        LongWritable count= new LongWritable(sum);
    context.write(key, count);
}
}
```

Key and Value Swapper class:
```java
package sayali.Assignment4Part6;

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

public class KeyValueSwappingMapper extends Mapper<Text, LongWritable, LongWritable,
Text> {

    public void map(Text key, LongWritable value, Context context) throws IOException,
InterruptedException {
        context.write(value, key);
    }
}
```

Chaining map reduce job:

```java
package sayali.Assignment4Part6;
import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.LongWritable;
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.SequenceFileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.SequenceFileOutputFormat;

public class top10IpAddress
{
```

```java
    public static void main( String[] args ) throws IOException, ClassNotFoundException,
InterruptedException

    {
        Configuration conf =new Configuration();
        Path out = new Path(args[1]);
         // Create a new Job
        Job job = Job.getInstance(conf, "Top10Ipaddress");
        job.setJarByClass(top10IpAddress.class);
        job.setJobName("myjob");
        job.setMapperClass(top10IpAddressMapper.class);
        job.setCombinerClass(top10IpAddressReducer.class);
        job.setReducerClass(top10IpAddressReducer.class);
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(LongWritable.class);
        job.setOutputFormatClass(SequenceFileOutputFormat.class);
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(out,"Out1"));
         if (!job.waitForCompletion(true)) {
         System.exit(1);
        }
        Job job2 = Job.getInstance(conf, "sort by frequency");
    job2.setJarByClass(top10IpAddress.class);
    job2.setMapperClass(KeyValueSwappingMapper.class);
    job2.setNumReduceTasks(1);
    job2.setSortComparatorClass(LongWritable.DecreasingComparator.class);
    job2.setOutputKeyClass(LongWritable.class);
    job2.setOutputValueClass(Text.class);
    job2.setInputFormatClass(SequenceFileInputFormat.class);
    FileInputFormat.addInputPath(job2, new Path(out, "Out1"));
    FileOutputFormat.setOutputPath(job2, new Path(out, "Out2"));
    if (!job2.waitForCompletion(true)) {
      System.exit(1);
    }
    }
}
```
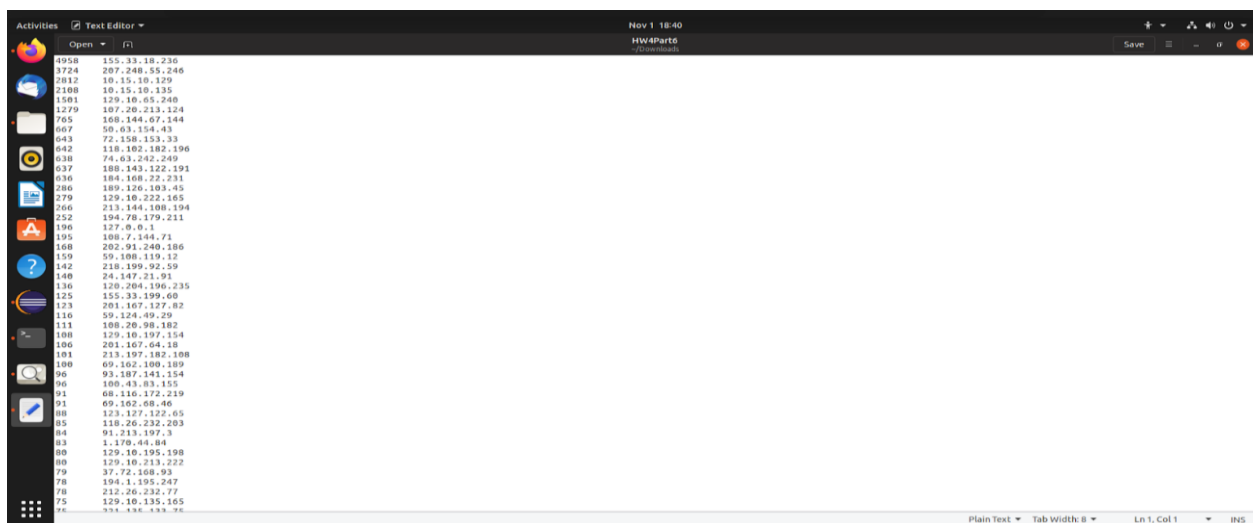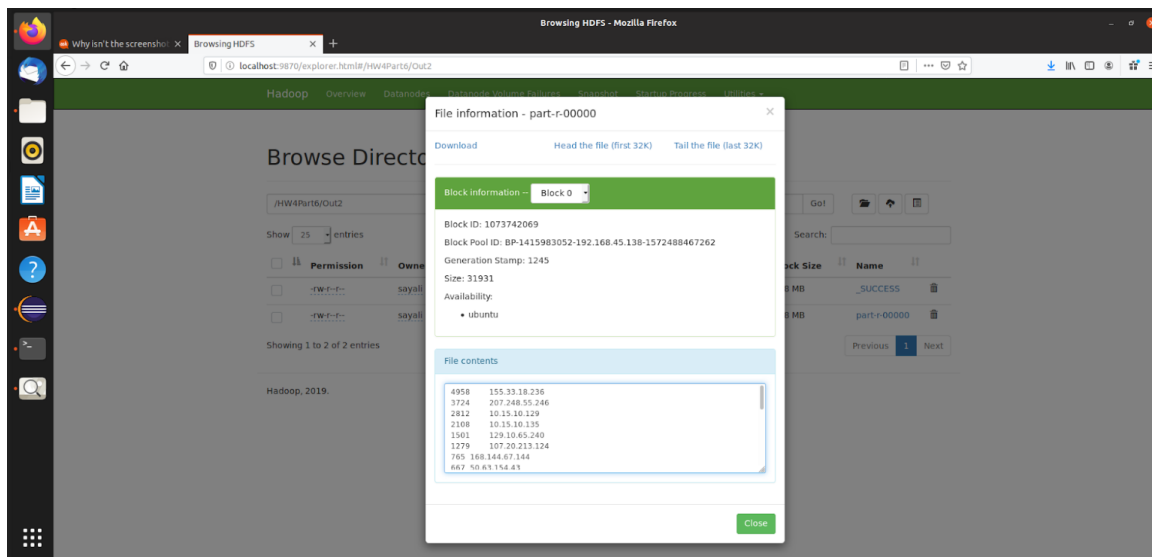
OUTPUT:
./hadoop jar /home/sayali/Desktop/HW4Part6.jar
sayali.Assignment4Part6.top10IpAddress /accessLog/access.csv /HW4Part6

Top 10 most visited IP addresses are:

./hadoop fs -cat /HW4Part6/Out2/part-r-00000 | head -n10

```
sayali@ubuntu:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop fs -cat /HW4Part6/Out2/part-r-00000 | head -n10
2019-11-01 18:48:16,915 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted
= false
4958    155.33.18.236
3724    207.248.55.246
2812    10.15.10.129
2108    10.15.10.135
1501    129.10.65.240
1279    107.20.213.124
765     168.144.67.144
667     50.63.154.43
643     72.158.153.33
642     118.102.182.196
cat: Unable to write to output stream.
sayali@ubuntu:/usr/local/bin/hadoop-3.2.1/bin$
```