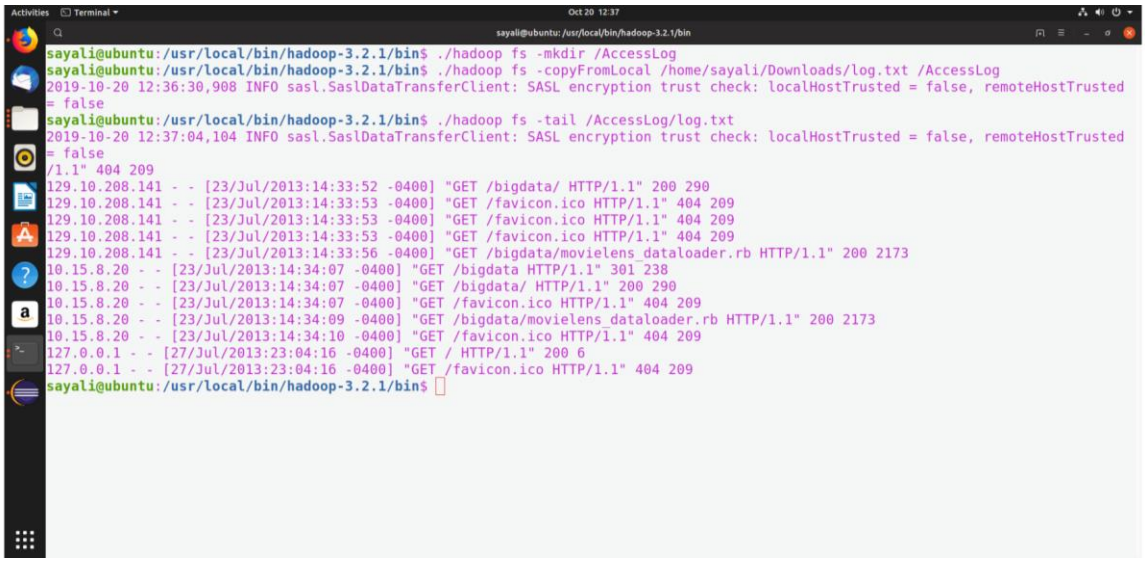
Map reduce on Access.csv file

1] MapReduce in Hadoop to find the number of times each IP accessed the website.

The access file looks like:

./hadoop fs -tail /AccessLog/access.csv



Mapper

```
package sayali.accesslog;
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 accessMapper extends Mapper<LongWritable, Text, Text, IntWritable>
        { @Override
        public void map(LongWritable key, Text value, Context context)throws IOException, InterruptedException
{           IntWritable one = new IntWritable(1);
                String line = value.toString();
                  List<String> items = Lists.newArrayList(Splitter.on(',').split(line));
                String ipaddress = items.get(0);
                String website=items.get(2);
                Text compositeKey = new Text(ipaddress+ "  " + website);
                context.write(compositeKey,one);  } }
```

Reducer :

```
package sayali.accesslog;
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;
public class accessReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
        public void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException,
InterruptedException
        {
                int sum = 0;
                for (IntWritable v : values)
                {
                sum += v.get();
}    IntWritable count= new IntWritable(sum);
    context.write(key, count);
} }
```

Main Class:

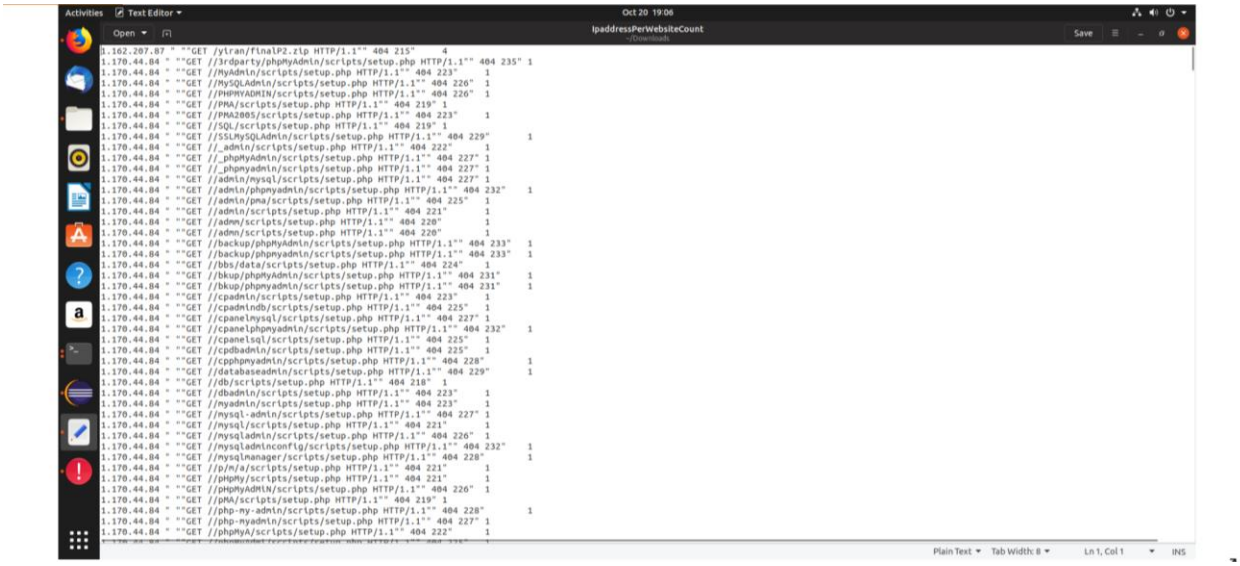
```
package sayali.accesslog;
import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;

public class App
{
   public static void main( String[] args ) throws IOException

   {
        Configuration conf =new Configuration();
         // Create a new Job
        Job job = Job.getInstance(conf, "Word count example");
        job.setJarByClass(App.class);
         // Specify various job-specific parameters
        job.setJobName("myjob");
        //set the mapper and reducer
        job.setMapperClass(accessMapper.class);
        job.setReducerClass(accessReducer.class);
        // set the format of mapper and reducer
        job.setInputFormatClass(TextInputFormat.class);
        job.setOutputFormatClass(TextOutputFormat.class);
        //set the key nd value format of the output
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);
```

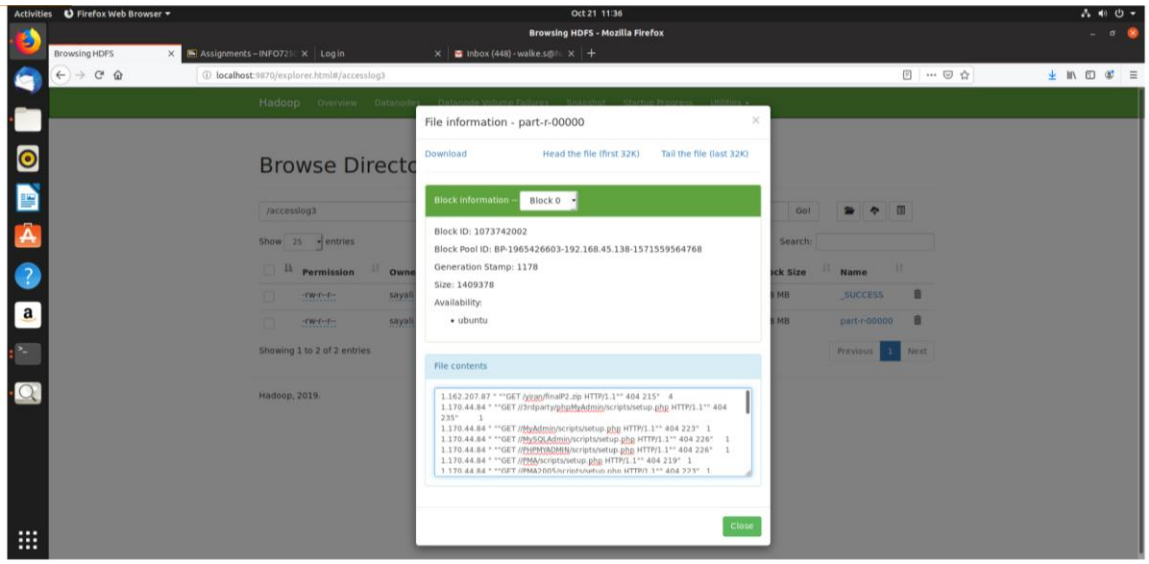
```
        //set the output and input format
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));
        // Submit the job, then poll for progress until the job is complete
        try {   job.waitForCompletion(true);
                } catch (ClassNotFoundException e) {

                        e.printStackTrace();
                } catch (InterruptedException e) {

                        e.printStackTrace();
                }
  } }
```

OUTPUT

./hadoop jar /home/sayali/Desktop/access.jar sayali.accesslog.App /AccessLog/access.csv /accesslog

The number of times each IP accessed the website.




Activities
Firefox Web Browser
Oct 21 11:36
Browsing HDFS - Mozilla Firefox
Browsing HDFS
Assignments – INFO721
Log in
Inbox (448) - walke.s@
localhost:9870/explorer.html#/accesslog3
Hadoop
Overview
Browse Directo
/accesslog3
Go!
Show 25 entries
Search:
Permission
Owne
Name
-rw-r--r--
sayali
_SUCCESS
part-r-00000
Showing 1 to 2 of 2 entries
Previous
1
Next
Hadoop, 2019.
File information - part-r-00000
Download
Head the file (first 32K)
Tail the file (last 32K)
Block information --
Block 0
Block ID: 1073742002
Block Pool ID: BP-1965426603-192.168.45.138-1571559564768
Generation Stamp: 1178
Size: 1409378
Availability:
ubuntu
File contents
1.162.207.87 - - "GET /vpn/finalP2.zip HTTP/1.1" 404 215" 4
1.170.44.84 - - "GET //3rdparty/phpMyAdmin/scripts/setup.php HTTP/1.1" 404 235" 1
1.170.44.84 - - "GET //MyAdmin/scripts/setup.php HTTP/1.1" 404 223" 1
1.170.44.84 - - "GET //MySQLAdmin/scripts/setup.php HTTP/1.1" 404 226" 1
1.170.44.84 - - "GET //PHPMYADMIN/scripts/setup.php HTTP/1.1" 404 226" 1
1.170.44.84 - - "GET //PMA/scripts/setup.php HTTP/1.1" 404 219" 1
Close