



SOMAIYA
VIDYAVIHAR UNIVERSITY

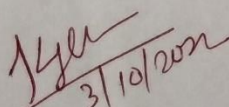
K J Somaiya Institute of Management

Department of Data Science & Technology

Certificate

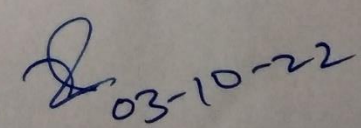
Masters in Computer Applications
Trimester IV (2022 – 23)

This is to certify that Mr. /Ms. Vaibhav Kumar Roll No. 19
of MCA, has satisfactorily completed the practical course “Big Data Analytics” prescribed by the College for the Partial fulfilment of the Degree by the Somaiya Vidyavihar University, during the academic year 2022-23.


Signature of the Faculty-Incharge

*Signature of the Programme
Coordinator*

3/10/2022
Date of Examination


Signature of the External Examiner/s

INDEX

Sr no	Topic Name	Date	Pg.NO.
1.	Write a program to find occurrence of any specific word	08-08-2022	4
2.	Average word count	12-08-2022	11
3.	Encrypt the email id from file	23-08-2022	14
4.	Multiple File Input	29-08-2022	17
5.	Web Log Analysis - Partitioner Assignment	01-09-2022	24
6.	Crime Data analysis using Pig	10-09-2022	29
7.	Web log + Cust Analysis using Pig	07-09-2022	34

Assignment 1 Big Data

Date :08-08-2022

Problem Statement:

Write a program to find occurrence of any specific word

This is the File created by me

```
import java.io.IOException;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable; //type casted to Int
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper.Context;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat; import
org.apache.hadoop.mapreduce.lib.output.FileOutputFormat; import
org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.mapreduce.Mapper; //mapper class
import org.apache.hadoop.mapreduce.Reducer;

//import CountWord.SumReducer;
//import CountWord.WordMapper;

public class CountWord {
    public static class WordMapper extends
    Mapper<LongWritable,Text,Text,IntWritable>{
        @Override
        public void map(LongWritable key, Text value, Context context)
        throws IOException, InterruptedException{
            String line = value.toString();

            for (String word : line.split("\\W+")){if
            (word.equals("hello")){
                word.toLowerCase();
                context.write(new Text(word), new IntWritable(1));
            }
        }
    }

    public static class SumReducer extends Reducer<Text, IntWritable,Text,
    IntWritable>{
```

```
@Override
public void reduce(Text key, Iterable<IntWritable> values, Context
context) throws IOException, InterruptedException{
    int wordCount = 0;

    for (IntWritable value: values){ wordCount
        += value.get();

    }
    context.write(key, new IntWritable(wordCount));
}

}

public static void main(String[] args) throws Exception{ if(args.length !=
2){
    System.out.printf("Usage: CountWord <input dir> <output dir>\n");
    System.exit(-1);

}
Job job = new Job();
job.setJarByClass(CountWord.class);
job.setJobName("CountWord");

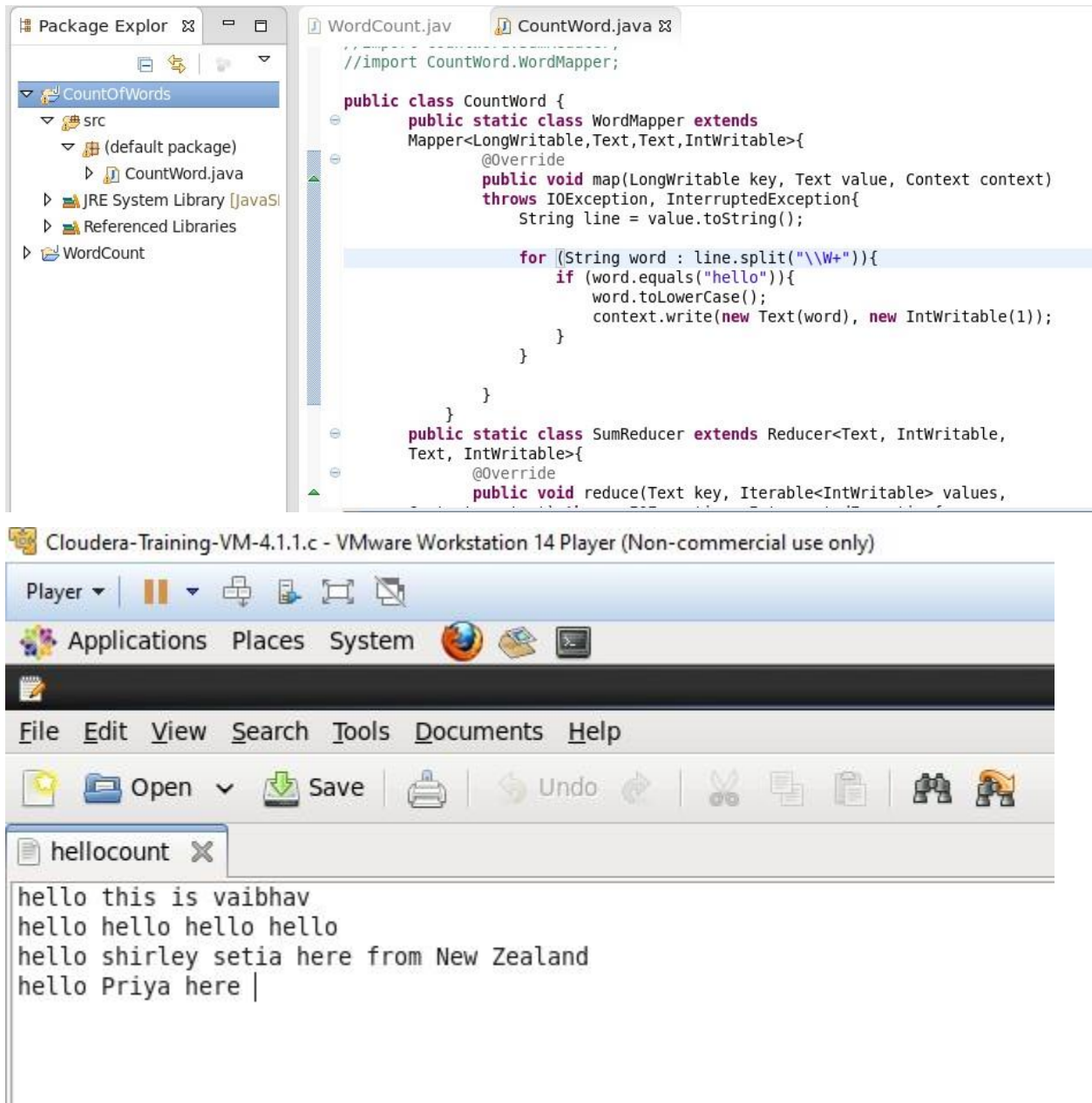
FileInputFormat.setInputPaths(job, new Path(args[0]));
FileOutputFormat.setOutputPath(job, new Path(args[1]));

job.setMapperClass(WordMapper.class);
job.setReducerClass(SumReducer.class);

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

boolean success = job.waitForCompletion(true);
System.exit(success ? 0 : 1);

}
}
```



Cloudera-Training-VM-4.1.1.c - VMware Workstation 14 Player (Non-commercial use only)

Player ▾ | [Icons: Full Screen, Zoom In, Zoom Out, Copy, Paste, Print]

Applications Places System [Icons: Firefox, File Manager, Terminal]

HDFS:/user/training

File Edit View History Bookmarks Tools Help

[HDFS:/user/training/vk/part-r-...] X [HDFS:/user/training/specifiko... X [Sign in - Google Account]

← [localhost.localdomain:50075/browseBlock.jsp?blockId=3852787711577530727&blockSize=

[Most Visited] [Hadoop NameNode] [Hadoop JobTracker] [Hadoop YARN] [Hue] [HBase]

File: [/user/training/specifiooutput/part-r-00000](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

```
hello 7
```

```
[training@localhost ~]$ hdfs dfs -copyFromLocal /home/training/Desktop/hellocount /user/training/vaibhav
[training@localhost ~]$ hadoop jar /home/training/specific.jar /user/training/vaibhav/hellocount /user/training/specificoutput
22/08/08 10:45:40 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for
the same.
22/08/08 10:45:40 INFO input.FileInputFormat: Total input paths to process : 1
22/08/08 10:45:40 WARN snappy.LoadSnappy: Snappy native library is available
22/08/08 10:45:40 INFO snappy.LoadSnappy: Snappy native library loaded
22/08/08 10:45:40 INFO mapred.JobClient: Running job: job_202208011504_0003
22/08/08 10:45:41 INFO mapred.JobClient: map 0% reduce 0%
22/08/08 10:45:45 INFO mapred.JobClient: map 100% reduce 0%
22/08/08 10:45:47 INFO mapred.JobClient: map 100% reduce 100%
22/08/08 10:45:47 INFO mapred.JobClient: Job complete: job_202208011504_0003
22/08/08 10:45:47 INFO mapred.JobClient: Counters: 32
22/08/08 10:45:47 INFO mapred.JobClient: File System Counters
22/08/08 10:45:47 INFO mapred.JobClient: FILE: Number of bytes read=90
22/08/08 10:45:47 INFO mapred.JobClient: FILE: Number of bytes written=361956
22/08/08 10:45:47 INFO mapred.JobClient: FILE: Number of read operations=0
22/08/08 10:45:47 INFO mapred.JobClient: FILE: Number of large read operations=0
22/08/08 10:45:47 INFO mapred.JobClient: FILE: Number of write operations=0
22/08/08 10:45:47 INFO mapred.JobClient: HDFS: Number of bytes read=226
22/08/08 10:45:47 INFO mapred.JobClient: HDFS: Number of bytes written=8
22/08/08 10:45:47 INFO mapred.JobClient: HDFS: Number of read operations=2

22/08/08 10:45:40 INFO input.FileInputFormat: Total input paths to process : 1
22/08/08 10:45:40 WARN snappy.LoadSnappy: Snappy native library is available
22/08/08 10:45:40 INFO snappy.LoadSnappy: Snappy native library loaded
22/08/08 10:45:40 INFO mapred.JobClient: Running job: job_202208011504_0003
22/08/08 10:45:41 INFO mapred.JobClient: map 0% reduce 0%
22/08/08 10:45:45 INFO mapred.JobClient: map 100% reduce 0%
22/08/08 10:45:47 INFO mapred.JobClient: map 100% reduce 100%
22/08/08 10:45:47 INFO mapred.JobClient: Job complete: job_202208011504_0003
22/08/08 10:45:47 INFO mapred.JobClient: Counters: 32
22/08/08 10:45:47 INFO mapred.JobClient: File System Counters
22/08/08 10:45:47 INFO mapred.JobClient: FILE: Number of bytes read=90
22/08/08 10:45:47 INFO mapred.JobClient: FILE: Number of bytes written=361956
22/08/08 10:45:47 INFO mapred.JobClient: FILE: Number of read operations=0
22/08/08 10:45:47 INFO mapred.JobClient: FILE: Number of large read operations=0
22/08/08 10:45:47 INFO mapred.JobClient: FILE: Number of write operations=0
22/08/08 10:45:47 INFO mapred.JobClient: HDFS: Number of bytes read=226
22/08/08 10:45:47 INFO mapred.JobClient: HDFS: Number of bytes written=8
22/08/08 10:45:47 INFO mapred.JobClient: HDFS: Number of read operations=2
22/08/08 10:45:47 INFO mapred.JobClient: HDFS: Number of large read operations=0
22/08/08 10:45:47 INFO mapred.JobClient: HDFS: Number of write operations=1
22/08/08 10:45:47 INFO mapred.JobClient: Job Counters
22/08/08 10:45:47 INFO mapred.JobClient: Launched map tasks=1
22/08/08 10:45:47 INFO mapred.JobClient: Launched reduce tasks=1
22/08/08 10:45:47 INFO mapred.JobClient: Data-local map tasks=1
22/08/08 10:45:47 INFO mapred.JobClient: Total time spent by all maps in occupied slots (ms)=3699
22/08/08 10:45:47 INFO mapred.JobClient: Total time spent by all reduces in occupied slots (ms)=2148
22/08/08 10:45:47 INFO mapred.JobClient: Total time spent by all maps waiting after reserving slots (ms)=0
22/08/08 10:45:47 INFO mapred.JobClient: Total time spent by all reduces waiting after reserving slots (ms)=0
22/08/08 10:45:47 INFO mapred.JobClient: Map-Reduce Framework
22/08/08 10:45:47 INFO mapred.JobClient: Map input records=4
```

Customer Data

```
public class CustomerData {
    public static class WordMapper extends
        Mapper<LongWritable,Text,Text,IntWritable>{
        @Override
        public void map(LongWritable key, Text value, Context context)
            throws IOException, InterruptedException{
            String line = value.toString();

            for (String word : line.split(",")){
                if (word.equals("Pilot")){
                    word.toLowerCase();
                    context.write(new Text(word), new IntWritable(1));
                }
            }
        }
    }

    public static class SumReducer extends Reducer<Text, IntWritable,Text,
        IntWritable>{
        @Override
        public void reduce(Text key, Iterable<IntWritable> values,Context
            context) throws IOException, InterruptedException{
            int wordCount = 0;

            for (IntWritable value: values){ wordCount
                += value.get();
            }
            context.write(key, new IntWritable(wordCount));
        }
    }

    public static void main(String[] args) throws Exception{ if(args.length !=
        2){
        System.out.printf("Usage: CountWord <input dir> <output dir>\n");
        System.exit(-1);
    }
    Job job = new Job();
    job.setJarByClass(CustomerData.class);
    job.setJobName("CustomerData");
```



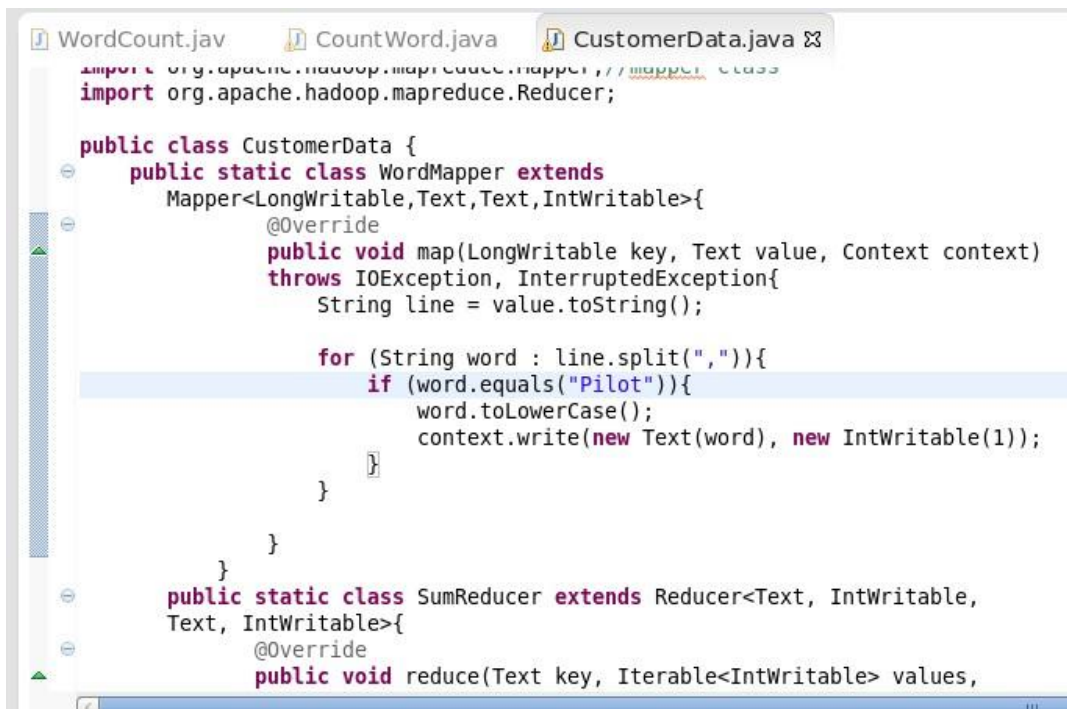
```
FileInputFormat.setInputPaths(job, new Path(args[0]));  
FileOutputFormat.setOutputPath(job, new Path(args[1]));
```

```
job.setMapperClass(WordMapper.class);  
job.setReducerClass(SumReducer.class);
```

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

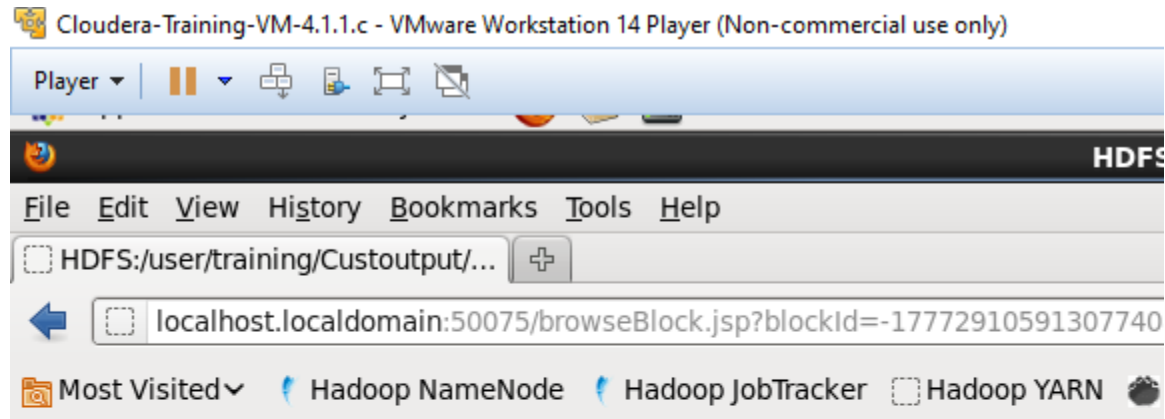
```
boolean success = job.waitForCompletion(true);  
System.exit(success ? 0 : 1);
```

```
}  
}
```



```
WordCount.java CountWord.java CustomerData.java ✖  
import org.apache.hadoop.mapreduce.Mapper; // mapper class  
import org.apache.hadoop.mapreduce.Reducer;  
  
public class CustomerData {  
    public static class WordMapper extends  
        Mapper<LongWritable, Text, Text, IntWritable> {  
        @Override  
        public void map(LongWritable key, Text value, Context context)  
            throws IOException, InterruptedException {  
            String line = value.toString();  
  
            for (String word : line.split(" ")) {  
                if (word.equals("Pilot")) {  
                    word.toLowerCase();  
                    context.write(new Text(word), new IntWritable(1));  
                }  
            }  
        }  
    }  
  
    public static class SumReducer extends Reducer<Text, IntWritable,  
        Text, IntWritable> {  
        @Override  
        public void reduce(Text key, Iterable<IntWritable> values,
```

```
[training@localhost ~]$ hdfs dfs -copyFromLocal /home/training/Desktop/custs /user/training/vaibhav
[training@localhost ~]$ hadoop jar /home/training/specificCustomer.jar /user/training/vaibhav/custs /user/training/Custoutput
22/08/08 11:05:21 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for
the same.
22/08/08 11:05:21 INFO input.FileInputFormat: Total input paths to process : 1
22/08/08 11:05:21 WARN snappy.LoadSnappy: Snappy native library is available
22/08/08 11:05:21 INFO snappy.LoadSnappy: Snappy native library loaded
22/08/08 11:05:22 INFO mapred.JobClient: Running job: job_202208011504_0004
22/08/08 11:05:23 INFO mapred.JobClient: map 0% reduce 0%
22/08/08 11:05:28 INFO mapred.JobClient: map 100% reduce 0%
22/08/08 11:05:32 INFO mapred.JobClient: map 100% reduce 100%
22/08/08 11:05:32 INFO mapred.JobClient: Job complete: job_202208011504_0004
22/08/08 11:05:32 INFO mapred.JobClient: Counters: 32
22/08/08 11:05:32 INFO mapred.JobClient:   File System Counters
22/08/08 11:05:32 INFO mapred.JobClient:     FILE: Number of bytes read=2514
22/08/08 11:05:32 INFO mapred.JobClient:     FILE: Number of bytes written=366804
22/08/08 11:05:32 INFO mapred.JobClient:     FILE: Number of read operations=0
22/08/08 11:05:32 INFO mapred.JobClient:     FILE: Number of large read operations=0
22/08/08 11:05:32 INFO mapred.JobClient:     FILE: Number of write operations=0
```



File: [/user/training/Custoutput/part-r-00000](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

Pilot 209

Assignment Big Data

Date:12-08-2022

Q) Write a program to find average word length

```
import java.io.IOException;
import java.util.*;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.DoubleWritable;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer.Context;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;

public class Average {

    public static class WordMapper extends Mapper <LongWritable, Text, Text, IntWritable>
    {
        @Override
        public void map(LongWritable key, Text value, Context context)
        throws IOException,
            InterruptedException{ Text firstLetter
            = new Text();
            IntWritable wordLength = new IntWritable();
            char[] a1 = null;
            String line = value.toString().toLowerCase();

            for(String word:line.split(" ")){

                if (word.length() > 0) {

                    firstLetter.set(String.valueOf(word.charAt(0)));

                    wordLength.set(word.length());
                    context.write(firstLetter, wordLength);
                }
            }
        }

        public static class SumReducer extends Reducer<Text,IntWritable,Text,IntWritable>{
            public void reduce(Text key,Iterable<IntWritable>values,Context context)
            throws IOException,InterruptedException{
                int sum =0;
                int count =0;
                int Average =0;
                for(IntWritable val:values)
                { sum += val.get();
```

```

        count = count+1;
    }
    Average = sum/count;

    context.write(key,new IntWritable(Average));
}

}

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

    if(args.length !=
    2){
        System.out.printf(
        "Usage: WordCount <input dir> <output dir> \n");

        System.exit(-1);
    }

    Job job = new Job();

    job.setJarByClass(Average.class);

    job.setJobName("Word Count");

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

    job.setMapperClass(WordMapper.class);
    job.setReducerClass(SumReducer.class);

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

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

```

eclipse



Videos

lib



apache-flume-1.4.0-
bin.tar.gz

Music



average011.jar

File: [/user/training/output10/part-r-00000](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

a	8
c	7
e	5
f	5
g	7
h	5
i	2
k	2
m	7
p	9
s	7
t	2
v	6
w	6
y	4

Assignment

Date : 23-08-2022

Problem Statement : Encrypt the email id from file

Explanation :

Implementing caesar cipher algorithm for this ,

Basically Caesar cipher or Shift Cipher is a Substitution cipher algorithm in which each letter of the plain text (message) is substituted with another letter. In this algorithm, each letter of the Plaintext is shifted a number of positions based on the Key provided.

Here i have provided the **key as 69**.

CODE:

```
import javax.crypto.Cipher;
import javax.crypto.KeyGenerator;
import javax.crypto.SecretKey;
import org.apache.hadoop.conf.Configuration; import
org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.Text; import
org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable; import
org.apache.hadoop.mapreduce.Mapper; import
org.apache.hadoop.mapreduce.Mapper; import
org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper.Context;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.util.GenericOptionsParser;
public class email {
    public static class emailencrypt extends Mapper<LongWritable,Text,Text,Text>
    {
        public void map(LongWritable key,Text value,Context con)throws
        IOException,InterruptedException
        {
            String strIncremented=new String();
            String x = "";
            String line = value.toString();
            for(int i=0;i<line.length();i++)
            {
                strIncremented+=(char)(line.charAt(i)+69);
            }
            con.write(new Text(x),new Text(strIncremented));
        }
    }
    }//end map
public static void main(String[] args) throws Exception{
```

```

Job job=new Job();
job.setJarByClass(email.class);
    job.setJobName("email"); FileInputFormat.setInputPaths(job,new
    Path(args[0])); FileOutputFormat.setOutputPath(job,new Path(args[1]));
job.setMapperClass(emailencrypt.class);
job.setOutputKeyClass(Text.class); job.setOutputValueClass(Text.class);
boolean success=job.waitForCompletion(true);
System.exit(success?0:1);    }}

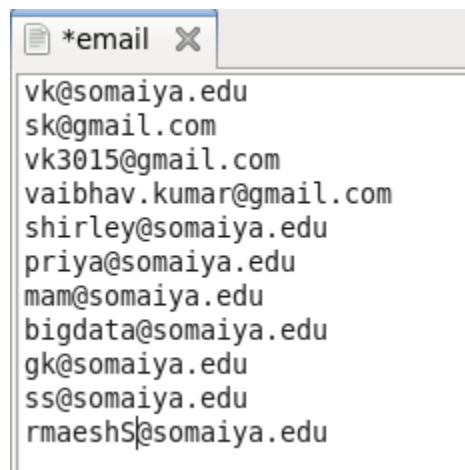
```

```

[training@localhost ~]$ hadoop jar /home/training/emailcry.jar /user/training/vaibhav/email /user/training/emailoutput
22/08/23 11:14:33 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for
the same.
22/08/23 11:14:33 INFO input.FileInputFormat: Total input paths to process : 1
22/08/23 11:14:33 WARN snappy.LoadSnappy: Snappy native library is available
22/08/23 11:14:33 INFO snappy.LoadSnappy: Snappy native library loaded
22/08/23 11:14:33 INFO mapred.JobClient: Running job: job_202208011504_0006
22/08/23 11:14:34 INFO mapred.JobClient: map 0% reduce 0%
22/08/23 11:14:38 INFO mapred.JobClient: map 100% reduce 0%
22/08/23 11:14:40 INFO mapred.JobClient: map 100% reduce 100%
22/08/23 11:14:40 INFO mapred.JobClient: Job complete: job_202208011504_0006
22/08/23 11:14:40 INFO mapred.JobClient: Counters: 32
22/08/23 11:14:40 INFO mapred.JobClient:   File System Counters
22/08/23 11:14:40 INFO mapred.JobClient:     FILE: Number of bytes read=398
22/08/23 11:14:40 INFO mapred.JobClient:     FILE: Number of bytes written=361496
22/08/23 11:14:40 INFO mapred.JobClient:     FILE: Number of read operations=0
22/08/23 11:14:40 INFO mapred.JobClient:     FILE: Number of large read operations=0
22/08/23 11:14:40 INFO mapred.JobClient:     FILE: Number of write operations=0
22/08/23 11:14:40 INFO mapred.JobClient:     HDFS: Number of bytes read=305
22/08/23 11:14:40 INFO mapred.JobClient:     HDFS: Number of bytes written=370
22/08/23 11:14:40 INFO mapred.JobClient:     HDFS: Number of read operations=2
22/08/23 11:14:40 INFO mapred.JobClient:     HDFS: Number of large read operations=0
22/08/23 11:14:40 INFO mapred.JobClient:     HDFS: Number of write operations=1
22/08/23 11:14:40 INFO mapred.JobClient:   Job Counters
22/08/23 11:14:40 INFO mapred.JobClient:     Launched map tasks=1
22/08/23 11:14:40 INFO mapred.JobClient:     Launched reduce tasks=1
22/08/23 11:14:40 INFO mapred.JobClient:     Data-local map tasks=1
22/08/23 11:14:40 INFO mapred.JobClient:     Total time spent by all maps in occupied slots (ms)=3540
22/08/23 11:14:40 INFO mapred.JobClient:     Total time spent by all reduces in occupied slots (ms)=2218
22/08/23 11:14:40 INFO mapred.JobClient:     Total time spent by all maps waiting after reserving slots (ms)=0

```

Input File :



*email X

```

vk@somaiya.edu
sk@gmail.com
vk3015@gmail.com
vaibhav.kumar@gmail.com
shirley@somaiya.edu
priya@somaiya.edu
mam@somaiya.edu
bigdata@somaiya.edu
gk@somaiya.edu
ss@somaiya.edu
rmaeshS@somaiya.edu

```

OUTPUT:

File: [/user/training/emailoutput/part-r-00000](#)

Goto :

[Go back to dir listing](#)

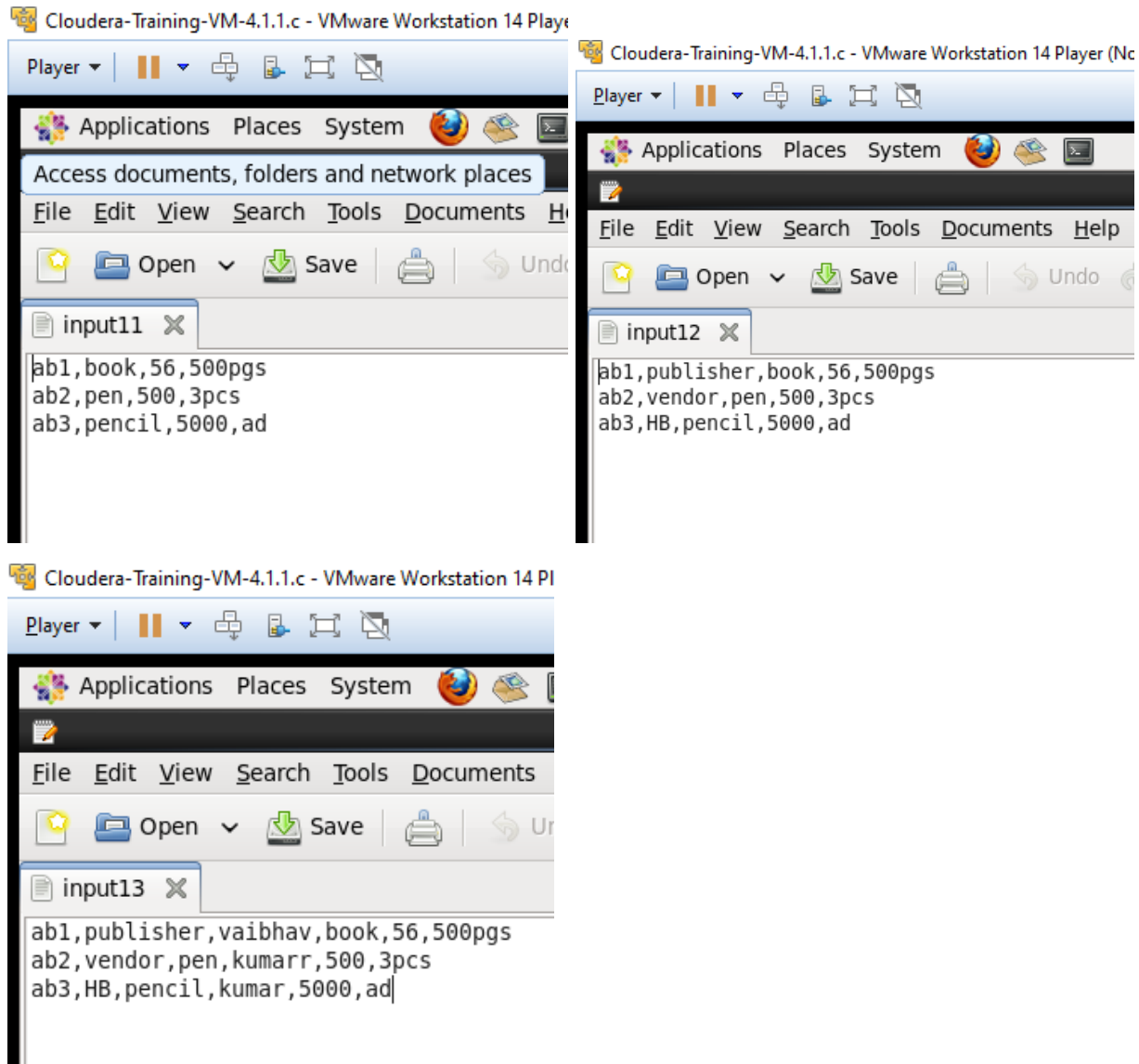
[Advanced view/download options](#)

$\begin{aligned}
& \gg \circ^{-2} \mid \mathbb{R}_4^3 \mid S \cong \mathbb{C}^0 \\
& \circ^{-2} \mid \mathbb{R} \mid S''^{-2} \\
& \gg \text{XUVZ}^{-2} \mid \mathbb{R} \mid S''^{-2} \\
& \gg \mid \mathbb{R} \mid S \mid \gg S^{\circ 2} \mid \cdot^{-2} \mid \mathbb{R} \mid S''^{-2} \\
& \mid \mathbb{R} \mid \pm \mathbb{R}_4^3 \mid \circ^{-2} \mid \mathbb{R}_4^3 \mid S \cong \mathbb{C}^0 \\
& \mu \cdot \mathbb{R}_4^3 \mid \circ^{-2} \mid \mathbb{R}_4^3 \mid S \cong \mathbb{C}^0 \\
& 2 \mid \circ^{-2} \mid \mathbb{R}_4^3 \mid S \cong \mathbb{C}^0 \\
& S \cong \mathbb{C}^0 \mid 1 \mid \circ^{-2} \mid \mathbb{R}_4^3 \mid S \cong \mathbb{C}^0 \\
& \circ^{-2} \mid \mathbb{R}_4^3 \mid S \cong \mathbb{C}^0 \\
& \circ^{-2} \mid \mathbb{R}_4^3 \mid S \cong \mathbb{C}^0 \\
& \cdot 2 \mid \mathbb{R} \mid \square \mid \circ^{-2} \mid \mathbb{R}_4^3 \mid S \cong \mathbb{C}^0
\end{aligned}$

ASSIGNMENT

Multiple File Input

Problem Statement: Create 2 or 3 input files on your own , in which the data is present in different format. Write a program to process these files using different map class and perform any one aggregate function like sum, max, min etc. on it.



CODE:

```
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.lib.input.MultipleInputs; import
org.apache.hadoop.mapreduce.lib.input.TextInputFormat; import
org.apache.hadoop.mapreduce.lib.output.FileOutputFormat; import
org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.io.LongWritable; import
org.apache.hadoop.mapreduce.Mapper;import
org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.util.GenericOptionsParser;

public class MultipleFileInp {
    public static class map1 extends Mapper <LongWritable, Text, Text,IntWritable>
    {
        @Override
        public void map(LongWritable key,Text value, Context con)
        throws IOException, InterruptedException{
            String line = value.toString();
            String[] line1=line.split(",");
            String category=line1[0];
            Text outputKey =new Text(category);int
            salary=Integer.parseInt(line1[2]);
            IntWritable outputValue=new IntWritable(salary);
            con.write(outputKey, outputValue);

        }
    }
    public static class map2 extends Mapper <LongWritable, Text, Text,IntWritable>
    {
        @Override
        public void map(LongWritable key,Text value, Context con)
        throws IOException, InterruptedException{
            String line = value.toString();
            String[] line1=line.split(",");
            String category=line1[0];
            Text outputKey =new Text(category);int
            salary=Integer.parseInt(line1[3]);
            IntWritable outputValue=new IntWritable(salary);
            con.write(outputKey, outputValue);

        }
    }
    public static class map3 extends Mapper <LongWritable, Text, Text,IntWritable>
```

```

{
    @Override
    public void map(LongWritable key, Text value, Context con)
    throws IOException, InterruptedException{
        String line = value.toString();
        String[] line1=line.split(",");
        String category=line1[0];
        Text outputKey =new Text(category);int
        salary=Integer.parseInt(line1[4]);
        IntWritable outputValue=new IntWritable(salary);
        con.write(outputKey, outputValue);

    }
}

public static class red extends Reducer<Text,IntWritable,Text,IntWritable>{
    public void reduce(Text category,Iterable<IntWritable> total_sal,Context con)throws
    IOException,InterruptedException{
        int sum =0;
        for(IntWritable value:total_sal){
            sum+=value.get();
        }
        con.write(category, new IntWritable(sum));
    }
}

}

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

    Configuration c = new Configuration();
    String[]files = new GenericOptionsParser(c,args).getRemainingArgs(); Path
    p1= new Path(files[0]);
    Path p2= new Path(files[1]);
    Path p3= new Path(files[2]);
    Path p4= new Path(files[3]);
    Job j =new Job(c,"multiple");
    j.setJarByClass(MultipleFileInp.class);
    j.setMapperClass(map1.class);
    j.setMapperClass(map2.class);
    j.setMapperClass(map3.class);
    j.setReducerClass(red.class);
    j.setMapOutputKeyClass(Text.class);
    j.setMapOutputValueClass(IntWritable.class);
    MultipleInputs.addInputPath(j, p1, TextInputFormat.class, map1.class);MultipleInputs.addInputPath(j, p2,
    TextInputFormat.class, map2.class);

```

```

MultipleInputs.addInputPath(j, p3, TextInputFormat.class, map3.class);
FileOutputFormat.setOutputPath(j, p4);
System.exit(j.waitForCompletion(true)?0:1);
}
}

```

```

[training@localhost ~]$ hadoop jar /home/training/multifile.jar /user/training/vaibhav/input11 /user/training/vaibhav/input12 /user/training/vaibhav/input13 /user/training/multioutput5
22/08/29 11:22:54 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
22/08/29 11:22:54 INFO input.FileInputFormat: Total input paths to process : 1
22/08/29 11:22:54 WARN snappy.LoadSnappy: Snappy native library is available
22/08/29 11:22:54 INFO snappy.LoadSnappy: Snappy native library loaded
22/08/29 11:22:54 INFO input.FileInputFormat: Total input paths to process : 1
22/08/29 11:22:54 INFO input.FileInputFormat: Total input paths to process : 1
22/08/29 11:22:54 INFO mapred.JobClient: Running job: job_202208011504_0011
22/08/29 11:22:55 INFO mapred.JobClient: map 0% reduce 0%
22/08/29 11:22:59 INFO mapred.JobClient: map 66% reduce 0%
22/08/29 11:23:01 INFO mapred.JobClient: map 100% reduce 0%
22/08/29 11:23:03 INFO mapred.JobClient: map 100% reduce 100%
22/08/29 11:23:03 INFO mapred.JobClient: Job complete: job_202208011504_0011
22/08/29 11:23:03 INFO mapred.JobClient: Counters: 32
22/08/29 11:23:03 INFO mapred.JobClient:   File System Counters
22/08/29 11:23:03 INFO mapred.JobClient:     FILE: Number of bytes read=96
22/08/29 11:23:03 INFO mapred.JobClient:     FILE: Number of bytes written=734280
22/08/29 11:23:03 INFO mapred.JobClient:     FILE: Number of read operations=0
22/08/29 11:23:03 INFO mapred.JobClient:     FILE: Number of large read operations=0
22/08/29 11:23:03 INFO mapred.JobClient:     FILE: Number of write operations=0
22/08/29 11:23:03 INFO mapred.JobClient:     HDFS: Number of bytes read=958
22/08/29 11:23:03 INFO mapred.JobClient:     HDFS: Number of bytes written=27
22/08/29 11:23:03 INFO mapred.JobClient:     HDFS: Number of read operations=6
22/08/29 11:23:03 INFO mapred.JobClient:     HDFS: Number of large read operations=0
22/08/29 11:23:03 INFO mapred.JobClient:     HDFS: Number of write operations=1
22/08/29 11:23:03 INFO mapred.JobClient:   Job Counters
22/08/29 11:23:03 INFO mapred.JobClient:     Launched map tasks=3
22/08/29 11:23:03 INFO mapred.JobClient:     Launched reduce tasks=1
22/08/29 11:23:03 INFO mapred.JobClient:     Data-local map tasks=3
22/08/29 11:23:03 INFO mapred.JobClient:     Total time spent by all maps in occupied slots (ms)=8012
22/08/29 11:23:03 INFO mapred.JobClient:     Total time spent by all reduces in occupied slots (ms)=3938
22/08/29 11:23:03 INFO mapred.JobClient:     Total time spent by all maps waiting after reserving slots (ms)=0
22/08/29 11:23:03 INFO mapred.JobClient:     Total time spent by all reduces waiting after reserving slots (ms)=0
22/08/29 11:23:03 INFO mapred.JobClient:   Map-Reduce Framework
22/08/29 11:23:03 INFO mapred.JobClient:     Map input records=9
22/08/29 11:23:03 INFO mapred.JobClient:     Map output records=9
22/08/29 11:23:03 INFO mapred.JobClient:     Map output bytes=72
22/08/29 11:23:03 INFO mapred.JobClient:     Input split bytes=732

```

File: [/user/training/multioutput5/part-r-00000](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

ab1	168
ab2	1500
ab3	15000

MAX

```
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.LongWritable; 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.MultipleInputs; import
org.apache.hadoop.mapreduce.lib.input.TextInputFormat; import
org.apache.hadoop.mapreduce.lib.output.FileOutputFormat; import
org.apache.hadoop.util.GenericOptionsParser;
public class MultipleFileInp {
    public static class map1 extends Mapper<LongWritable,Text,Text,IntWritable>
    {
        public void map(LongWritable key, Text value, Context con) throws IOException,
        InterruptedException
        {
            String line = value.toString();
            String[] line1 = line.split(",");
            String name = line1[0];
            Text outputKey = new Text(name);int
            cost = Integer.parseInt(line1[2]);
            IntWritable outputValue = new IntWritable(cost);
            con.write(outputKey, outputValue);
        } }
    public static class Map2 extends Mapper<LongWritable, Text, Text, IntWritable>
    {
        public void map(LongWritable key, Text value, Context con) throws IOException,
        InterruptedException
        {
            String line = value.toString();
            String[] line1 = line.split(",");
            String name = line1[0];
            Text outputKey = new Text(name);int
            cost = Integer.parseInt(line1[3]);
            IntWritable outputValue = new IntWritable(cost);
            con.write(outputKey, outputValue);
        } }
    public static class Map3 extends Mapper<LongWritable, Text, Text, IntWritable>
```

```

{

    public void map(LongWritable key, Text value, Context con) throws IOException,
    InterruptedException
    {
        String line = value.toString();
        String[] line1 = line.split(",");
        String name = line1[0];
        Text outputKey = new Text(name);int
        cost = Integer.parseInt(line1[4]);
        IntWritable outputValue = new IntWritable(cost);
        con.write(outputKey, outputValue);
    }
}

public static class Red extends Reducer<Text, IntWritable, Text,IntWritable>
{
    public void reduce(Text name, Iterable<IntWritable>total_cost, Context con) throws IOException,
    InterruptedException
    {
        int sum = 0;
        int max = 0;
        for (IntWritable value : total_cost) {
            if (value.get() > max) {
                max = value.get();
            }
        }
        con.write(name,new IntWritable(sum));
    }
}

public static void main(String[] args) throws Exception{
    Configuration c = new Configuration();
    GenericOptionsParser parser = new GenericOptionsParser(c,args);
    String[] files = parser.getRemainingArgs();
    Path p1 = new Path(files[0]);
    Path p2 = new Path(files[1]);
    Path p3 = new Path(files[2]);
    Path p4 = new Path(files[3]);
    Job j = new Job(c,"multiple");
    j.setJarByClass(MultipleFileInp .class);
    j.setMapperClass(map1.class);
    j.setMapperClass(Map2.class);
    j.setMapperClass(Map3.class);
    j.setReducerClass(Red.class);
    j.setOutputKeyClass(Text.class);

```

```

j.setOutputValueClass(IntWritable.class);
MultipleInputs.addInputPath(j,p1,TextInputFormat.class,map1.class);
MultipleInputs.addInputPath(j, p2, TextInputFormat.class,Map2.class);
MultipleInputs.addInputPath(j, p3, TextInputFormat.class,Map3.class);
FileOutputFormat.setOutputPath(j,p4);
System.exit(j.waitForCompletion(true)?0:1);} }

```

```

[training@localhost ~]$ hadoop jar /home/training/multifilemax.jar /user/training/vaibhav/input11 /user/training/vaibhav/input12 /user/training/vaibhav/input13 /user/training/multimax
22/08/29 11:42:39 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
22/08/29 11:42:39 INFO input.FileInputFormat: Total input paths to process : 1
22/08/29 11:42:39 WARN snappy.LoadSnappy: Snappy native library is available
22/08/29 11:42:39 INFO snappy.LoadSnappy: Snappy native library loaded
22/08/29 11:42:39 INFO input.FileInputFormat: Total input paths to process : 1
22/08/29 11:42:39 INFO input.FileInputFormat: Total input paths to process : 1
22/08/29 11:42:39 INFO mapred.JobClient: Running job: job_202208011504_0012
22/08/29 11:42:40 INFO mapred.JobClient: map 0% reduce 0%
22/08/29 11:42:44 INFO mapred.JobClient: map 66% reduce 0%
22/08/29 11:42:46 INFO mapred.JobClient: map 100% reduce 0%
22/08/29 11:42:48 INFO mapred.JobClient: map 100% reduce 100%
22/08/29 11:42:48 INFO mapred.JobClient: Job complete: job_202208011504_0012
22/08/29 11:42:48 INFO mapred.JobClient: Counters: 32
22/08/29 11:42:48 INFO mapred.JobClient: File System Counters
22/08/29 11:42:48 INFO mapred.JobClient: FILE: Number of bytes read=96
22/08/29 11:42:48 INFO mapred.JobClient: FILE: Number of bytes written=734240
22/08/29 11:42:48 INFO mapred.JobClient: FILE: Number of read operations=0
22/08/29 11:42:48 INFO mapred.JobClient: FILE: Number of large read operations=0
22/08/29 11:42:48 INFO mapred.JobClient: FILE: Number of write operations=0
22/08/29 11:42:48 INFO mapred.JobClient: HDFS: Number of bytes read=958
22/08/29 11:42:48 INFO mapred.JobClient: HDFS: Number of bytes written=18
22/08/29 11:42:48 INFO mapred.JobClient: HDFS: Number of read operations=6
22/08/29 11:42:48 INFO mapred.JobClient: HDFS: Number of large read operations=0
22/08/29 11:42:48 INFO mapred.JobClient: HDFS: Number of write operations=1
22/08/29 11:42:48 INFO mapred.JobClient: Job Counters
22/08/29 11:42:48 INFO mapred.JobClient: Launched map tasks=3
22/08/29 11:42:48 INFO mapred.JobClient: Launched reduce tasks=1
22/08/29 11:42:48 INFO mapred.JobClient: Data-local map tasks=3
22/08/29 11:42:48 INFO mapred.JobClient: Total time spent by all maps in occupied slots (ms)=7836

```

[Go back to dir listing](#)

[Advanced view/download options](#)

Ab1	570
Ab1	1670
ab2	600
ab2	1500
ab3	5500
ab3	16100

Web Log Analysis - Partitioner Assignment

Date : 01-09-22

For Odd Roll No:

Find Out the hit count per month

CODE:

```
import java.io.*;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.*;
import org.apache.hadoop.conf.*;
import org.apache.hadoop.fs.*;
import org.apache.hadoop.mapreduce.lib.input.*;
import org.apache.hadoop.mapreduce.lib.output.*;
import org.apache.hadoop.util.*;

public class multiFilesPart extends Configured implements Tool{
//map class
public static class Map extends Mapper<LongWritable, Text, Text, IntWritable> {
@Override public void map (LongWritable key, Text value, Context context) throws
IOException, InterruptedException {
String line = value.toString();
String[] word =line.split("/") ; String
code=word[1];
context.write(new Text (code), new IntWritable (1));}}
//Reducer class
public static class SumReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
@Override public void reduce (Text key, Iterable<IntWritable> values, Context context) throws
IOException, InterruptedException {
int wordCount = 0;
for (IntWritable value : values) {
wordCount += value.get();
}
context.write(key, new IntWritable (wordCount));}}
//Partitioner class
public static class CaderPartitioner extends Partitioner<Text,IntWritable>{ public int
getPartition ( Text key , IntWritable value , int numReduceTasks )
{
if ( numReduceTasks == 0 )
{
return 0 ;
}
if ( key.toString().equals("Jan"))
{
return 1 % numReduceTasks ;
}
else if (key.toString().equals("Feb"))
```



```
{
    return 2% numReduceTasks;
}
else if ( key.toString().equals("Mar") )
{
    return 3% numReduceTasks;
}
else if ( key.toString().equals("Apr") )
{
    return 4% numReduceTasks;
}
else if ( key.toString().equals("May") )
{
    return 5% numReduceTasks;
}
else if ( key.toString().equals("Jun") )
{
    return 6% numReduceTasks;
}
else if ( key.toString().equals("Jul") )
{
    return 7% numReduceTasks;
}
else if ( key.toString().equals("Aug") )
{
    return 8% numReduceTasks;
}
else if ( key.toString().equals("Sep") )
{
    return 9% numReduceTasks;
}
else if ( key.toString().equals("Oct") )
{
    return 10% numReduceTasks;
}
else if ( key.toString().equals("Nov") )
{
    return 11% numReduceTasks;
}
else if ( key.toString().equals("Dec") )
{
    return 12% numReduceTasks;
}
else{
```

```

    return 0;}}
public static void main(String[] args) throws Exception{
int res=ToolRunner.run(new Configuration(),new multiFilesPart(), args);
System.exit(0);}@Override
public int run(String[] arg) throws Exception {
Configuration conf = getConf() ;
    Job job = new Job (conf ,"Partitioner2" ) ;
    job.setJarByClass(multiFilesPart.class ) ;
    FileInputFormat.setInputPaths (job ,new Path ( arg [ 0 ]));
    FileOutputFormat.setOutputPath (job ,new Path ( arg [ 1 ]));
    job.setMapperClass(Map.class ) ; job.setMapOutputKeyClass (
    Text.class ) ; job.setMapOutputValueClass (IntWritable.class ) ;
    // set partitioner statement job.setPartitionerClass
    (CaderPartitioner.class ) ;
    job.setReducerClass(SumReducer.class);
    job.setNumReduceTasks(13) ;
    job.setInputFormatClass (TextInputFormat.class ) ;
    job.setOutputFormatClass(TextOutputFormat.class ) ;
    job.setOutputKeyClass ( Text.class ) ;
    job.setOutputValueClass ( Text.class ) ;
    System.exit(job.waitForCompletion ( true) ? 0 : 1); return 0 ;
}

```

File: [/user/training/vaibhav/access_log](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

[View Next chunk](#)

```

10.223.157.186 - - [15/Jul/2009:14:58:59 -0700] "GET / HTTP/1.1" 403 202
10.223.157.186 - - [15/Jul/2009:14:58:59 -0700] "GET /favicon.ico HTTP/1.1" 404 209
10.223.157.186 - - [15/Jul/2009:15:50:35 -0700] "GET / HTTP/1.1" 200 9157
10.223.157.186 - - [15/Jul/2009:15:50:35 -0700] "GET /assets/js/lowpro.js HTTP/1.1" 200 10469
10.223.157.186 - - [15/Jul/2009:15:50:35 -0700] "GET /assets/css/reset.css HTTP/1.1" 200 1014
10.223.157.186 - - [15/Jul/2009:15:50:35 -0700] "GET /assets/css/960.css HTTP/1.1" 200 6206
10.223.157.186 - - [15/Jul/2009:15:50:35 -0700] "GET /assets/css/the-associates.css HTTP/1.1" 200 15779
10.223.157.186 - - [15/Jul/2009:15:50:35 -0700] "GET /assets/js/the-associates.js HTTP/1.1" 200 4492
10.223.157.186 - - [15/Jul/2009:15:50:35 -0700] "GET /assets/js/lightbox.js HTTP/1.1" 200 25960
10.223.157.186 - - [15/Jul/2009:15:50:36 -0700] "GET /assets/img/search-button.gif HTTP/1.1" 200 168
10.223.157.186 - - [15/Jul/2009:15:50:36 -0700] "GET /assets/img/dummy/secondary-news-3.jpg HTTP/1.1" 200 5604
10.223.157.186 - - [15/Jul/2009:15:50:36 -0700] "GET /assets/img/dummy/primary-news-1.jpg HTTP/1.1" 200 10556
10.223.157.186 - - [15/Jul/2009:15:50:36 -0700] "GET /assets/img/dummy/primary-news-2.jpg HTTP/1.1" 200 9925
10.223.157.186 - - [15/Jul/2009:15:50:36 -0700] "GET /assets/img/closelabel.gif HTTP/1.1" 200 979
10.223.157.186 - - [15/Jul/2009:15:50:36 -0700] "GET /assets/img/home-logo.png HTTP/1.1" 200 3892
10.223.157.186 - - [15/Jul/2009:15:50:36 -0700] "GET /assets/img/dummy/secondary-news-2.jpg HTTP/1.1" 200 5397
10.223.157.186 - - [15/Jul/2009:15:50:36 -0700] "GET /assets/img/loading.gif HTTP/1.1" 200 2767
10.223.157.186 - - [15/Jul/2009:15:50:36 -0700] "GET /assets/img/dummy/secondary-news-4.jpg HTTP/1.1" 200 5766
10.223.157.186 - - [15/Jul/2009:15:50:36 -0700] "GET /assets/img/home-media-block-placeholder.jpg HTTP/1.1" 200 68831
10.223.157.186 - - [15/Jul/2009:15:50:37 -0700] "GET /assets/img/dummy/secondary-news-1.jpg HTTP/1.1" 200 5766
10.223.157.186 - - [15/Jul/2009:15:50:37 -0700] "GET /assets/swf/home-media-block.swf HTTP/1.1" 200 123884
10.223.157.186 - - [15/Jul/2009:15:50:51 -0700] "GET / HTTP/1.1" 200 9157
10.223.157.186 - - [15/Jul/2009:15:50:51 -0700] "GET /assets/css/960.css HTTP/1.1" 304 -
10.223.157.186 - - [15/Jul/2009:15:50:51 -0700] "GET /assets/css/the-associates.css HTTP/1.1" 304 -
10.223.157.186 - - [15/Jul/2009:15:50:51 -0700] "GET /assets/js/lowpro.js HTTP/1.1" 304 -
10.223.157.186 - - [15/Jul/2009:15:50:51 -0700] "GET /assets/js/lightbox.js HTTP/1.1" 304 -

```

```

[training@localhost ~]$ hadoop jar /home/training/part.jar /user/training/vaibha
v/access_log /user/training/partoutput
22/09/01 11:20:56 WARN mapred.JobClient: Use GenericOptionsParser for parsing th
e arguments. Applications should implement Tool for the same.
22/09/01 11:20:56 INFO input.FileInputFormat: Total input paths to process : 1
22/09/01 11:20:56 WARN snappy.LoadSnappy: Snappy native library is available
22/09/01 11:20:56 INFO snappy.LoadSnappy: Snappy native library loaded
22/09/01 11:20:56 INFO mapred.JobClient: Running job: job_202208300919_0001
22/09/01 11:20:57 INFO mapred.JobClient: map 0% reduce 0%
22/09/01 11:21:07 INFO mapred.JobClient: map 18% reduce 0%
22/09/01 11:21:10 INFO mapred.JobClient: map 25% reduce 0%
22/09/01 11:21:19 INFO mapred.JobClient: map 33% reduce 1%
22/09/01 11:21:20 INFO mapred.JobClient: map 46% reduce 1%
22/09/01 11:21:22 INFO mapred.JobClient: map 50% reduce 1%
22/09/01 11:21:25 INFO mapred.JobClient: map 50% reduce 2%
22/09/01 11:21:29 INFO mapred.JobClient: map 62% reduce 2%
22/09/01 11:21:30 INFO mapred.JobClient: map 73% reduce 2%
22/09/01 11:21:31 INFO mapred.JobClient: map 75% reduce 2%
22/09/01 11:21:32 INFO mapred.JobClient: map 75% reduce 3%
22/09/01 11:21:37 INFO mapred.JobClient: map 87% reduce 3%
22/09/01 11:21:38 INFO mapred.JobClient: map 100% reduce 3%
22/09/01 11:21:40 INFO mapred.JobClient: map 100% reduce 15%
22/09/01 11:21:44 INFO mapred.JobClient: map 100% reduce 30%
22/09/01 11:21:47 INFO mapred.JobClient: map 100% reduce 38%
22/09/01 11:21:48 INFO mapred.JobClient: map 100% reduce 46%
22/09/01 11:21:51 INFO mapred.JobClient: map 100% reduce 61%
22/09/01 11:21:55 INFO mapred.JobClient: map 100% reduce 69%
22/09/01 11:21:56 INFO mapred.JobClient: map 100% reduce 76%
22/09/01 11:21:59 INFO mapred.JobClient: map 100% reduce 84%
22/09/01 11:22:00 INFO mapred.JobClient: map 100% reduce 92%
22/09/01 11:22:02 INFO mapred.JobClient: map 100% reduce 100%
22/09/01 11:22:02 INFO mapred.JobClient: Job complete: job_202208300919_0001
22/09/01 11:22:02 INFO mapred.JobClient: Counters: 32
22/09/01 11:22:02 INFO mapred.JobClient: File System Counters
22/09/01 11:22:02 INFO mapred.JobClient: FILE: Number of bytes read=90232994
22/09/01 11:22:02 INFO mapred.JobClient: FILE: Number of bytes written=138205541
22/09/01 11:22:02 INFO mapred.JobClient: FILE: Number of read operations=0
22/09/01 11:22:02 INFO mapred.JobClient: FILE: Number of large read operations=0
22/09/01 11:22:02 INFO mapred.JobClient: FILE: Number of write operations=0
}

```

Cloudera-Training-VM-4.1.1.c - VMware Workstation 14 Player (Non-commercial use only)

Player ▾ | Applications Places System |

HDFS:/user/training/partou

File Edit View History Bookmarks Tools Help

HDFS:/user/training/partoutput

localhost.localdomain:50075/browseDirectory.jsp?dir=/user/training/partoutput&namenodeInfoPort=50070&nna

Most Visited ▾ | Hadoop NameNode | Hadoop JobTracker | Hadoop YARN | Hue | HBase Master

Contents of directory /user/training/partoutput

Goto :

[Go to parent directory](#)

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
SUCCESS	file	0 KB	1	64 MB	2022-09-01 11:22	rw-r--r--	training	supergroup
_logs	dir				2022-09-01 11:20	rw-r--r--	training	supergroup
part-r-00000	file	0.05 KB	1	64 MB	2022-09-01 11:21	rw-r--r--	training	supergroup
part-r-00001	file	0.01 KB	1	64 MB	2022-09-01 11:21	rw-r--r--	training	supergroup
part-r-00002	file	0.01 KB	1	64 MB	2022-09-01 11:21	rw-r--r--	training	supergroup
part-r-00003	file	0.01 KB	1	64 MB	2022-09-01 11:21	rw-r--r--	training	supergroup
part-r-00004	file	0.01 KB	1	64 MB	2022-09-01 11:21	rw-r--r--	training	supergroup
part-r-00005	file	0.01 KB	1	64 MB	2022-09-01 11:21	rw-r--r--	training	supergroup
part-r-00006	file	0.01 KB	1	64 MB	2022-09-01 11:21	rw-r--r--	training	supergroup
part-r-00007	file	0.01 KB	1	64 MB	2022-09-01 11:21	rw-r--r--	training	supergroup
part-r-00008	file	0.01 KB	1	64 MB	2022-09-01 11:21	rw-r--r--	training	supergroup
part-r-00009	file	0.01 KB	1	64 MB	2022-09-01 11:21	rw-r--r--	training	supergroup
part-r-00010	file	0.01 KB	1	64 MB	2022-09-01 11:21	rw-r--r--	training	supergroup
part-r-00011	file	0.01 KB	1	64 MB	2022-09-01 11:21	rw-r--r--	training	supergroup
part-r-00012	file	0.01 KB	1	64 MB	2022-09-01 11:21	rw-r--r--	training	supergroup

File: [/user/training/partoutput/part-r-00001](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

Jan 273096

File: [/user/training/partoutput/part-r-00011](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

Nov 421626

File: [/user/training/partoutput/part-r-00009](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

Sep 430529

File: [/user/training/partoutput/part-r-00005](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

May 339551

Assignment

Crime Data analysis using Pig

1. To find the total number of crimes which occurred in all the states in the year 2006.

L = LOAD '/user/training/vaibhav/crimeyear1.csv' using PigStorage(',') AS

(State:chararray,crimehead:chararray,y1:double,y2:double,y3:double,y4:double,y5:double,y6:double,y7:double,y8:double,y9:double,y10:double,y11:double,y12:double);

```
grunt> L = LOAD '/user/training/vaibhav/crimeyear1.csv' using PigStorage(',')
AS (State:chararray,crimehead:chararray,y1:double,y2:double,y3:double,y4:double,y5:double,y6:double,y7:double,y8:double,y9:double,y10:double,y11:double,y12:double);
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - dfs.permissions.supergroup is deprecated. Instead, use dfs.permissions.superusergroup
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - dfs.max.objects is deprecated. Instead, use dfs.namenode.max.objects
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - dfs.replication.interval is deprecated. Instead, use dfs.namenode.replication.interval
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - dfs.data.dir is deprecated. Instead, use dfs.datanode.data.dir
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - dfs.access.time.precision is deprecated. Instead, use dfs.namenode.access.time.precision
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - dfs.replication.min is deprecated. Instead, use dfs.namenode.replication.min
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - fs.checkpoint.dir is deprecated. Instead, use dfs.namenode.checkpoint.dir
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - dfs.http.address is deprecated. Instead, use dfs.namenode.http.address
2022-09-08 11:02:03,940 [main] WARN org.apache.hadoop.conf.Configuration - dfs.replication.considerLoad is deprecated. Instead, use dfs.namenode.replication.considerLoad
```

grunt> G = Group L all;

grunt> S = FOREACH G GENERATE SUM(L.y6);

grunt> describe S;S:

{double}

```
grunt> G = Group L all;
grunt> S = FOREACH G GENERATE SUM(L.y6);
grunt> describe S;
S: {double}
grunt> dump S;
2022-09-08 11:03:47,895 [main] INFO org.apache.
```

```
2022-09-08 11:04:01,082 [mai
2022-09-08 11:04:01,082 [mai
(559646.0)
```

2. Calculate the number of females (age between 18-30 years) who were victims in different crimes in different states.

```
L = LOAD '/user/training/vaibhav/crimeage1.csv' using PigStorage(',') AS
(state:chararray,crime:chararray,M18:int,F18:int,m1830:int,f1830:int,m3045:int,f3045:int,m4560:i
nt,f4560:int,m60:int,f60:int,tm:int,tf:int,t:int);
```

```
grunt> L = LOAD '/user/training/vaibhav/crimeage1.csv' using PigStorage(',') AS (state:chararray,
crime:chararray,M18:int,F18:int,m1830:int,f1830:int,m3045:int,f3045:int,m4560:int,f4560:int,m60:i
nt,f60:int,tm:int,tf:int,t:int);
2022-09-08 11:43:59,953 [main] WARN org.apache.hadoop.conf.Configuration - io.bytes.per.checksum
is deprecated. Instead, use dfs.bytes-per-checksum
```

```
C = FOREACH L generate crime,f1830;P =
GROUP C ALL;
R = FOREACH P GENERATE SUM(C.$1);
```

```
grunt> C = FOREACH L generate crime,f1830;
grunt> P = GROUP C ALL;
grunt> describe C;
C: {crime: chararray,f1830: int}
grunt> describe P;
P: {group: chararray,C: {(crime: chararray,f1830: int)}}
grunt> R = FOREACH P GENERATE SUM(C.$1);
grunt> describe R;
R: {long}
```

```
2022-09-08 11:45:52,
2022-09-08 11:45:52,
2022-09-08 11:45:52,
(86071)
grunt>
```

3. Which state have highest rape cases across all years.

YC = FILTER D BY (crime matches '.*(RAPE).*');dump

YC;

CV = FOREACH YC generate state,crime,t;

dump CV;

GC = GROUP CV BY state;

```
grunt> D = LOAD '/user/training/kirti/crimeagel.csv' using PigStorage(',') AS
0:int,tm:int,tf:int,t:int);
grunt> CC = FOREACH D generate state,crime,t;
grunt> dump CC;
```

```
2022-09-08 12:00:41,270 [main] INFO org.apache.p
(ANDHRA PRADESH,RAPE (SECTION 376 IPC),1664)
(ARUNACHAL PRADESH,RAPE (SECTION 376 IPC),47)
(ASSAM,RAPE (SECTION 376 IPC),1626)
(BIHAR,RAPE (SECTION 376 IPC),1327)
(CHHATTISGARH,RAPE (SECTION 376 IPC),1214)
(GOA,RAPE (SECTION 376 IPC),61)
(GUJARAT,RAPE (SECTION 376 IPC),647)
(HARYANA,RAPE (SECTION 376 IPC),940)
(HIMACHAL PRADESH,RAPE (SECTION 376 IPC),259)
(JAMMU & KASHMIR,RAPE (SECTION 376 IPC),388)
(JHARKHAND,RAPE (SECTION 376 IPC),780)
(KARNATAKA,RAPE (SECTION 376 IPC),842)
(KERALA,RAPE (SECTION 376 IPC),1259)
(MADHYA PRADESH,RAPE (SECTION 376 IPC),4822)
(MAHARASHTRA,RAPE (SECTION 376 IPC),2591)
(MANIPUR,RAPE (SECTION 376 IPC),46)
(MEGHALAYA,RAPE (SECTION 376 IPC),182)
(MIZORAM,RAPE (SECTION 376 IPC),122)
(NAGALAND,RAPE (SECTION 376 IPC),26)
(ODISHA,RAPE (SECTION 376 IPC),1666)
(PUNJAB,RAPE (SECTION 376 IPC),895)
(RAJASTHAN,RAPE (SECTION 376 IPC),1807)
(SIKKIM,RAPE (SECTION 376 IPC),29)
(TAMIL NADU,RAPE (SECTION 376 IPC),962)
(TRIPURA,RAPE (SECTION 376 IPC),202)
(UTTAR PRADESH,RAPE (SECTION 376 IPC),3593)
(UTTARAKHAND,RAPE (SECTION 376 IPC),184)
```



```

(Y DOCUMENT, TOTAL CRIMES AGAINST WOMEN, 110)
grunt> YC = FILTER CB BY (crime matches '.*(RAPE).*');
grunt> dump YC;

grunt> GC = GROUP CV BY state;
grunt> dump GC;
2022-09-08 12:08:19,863 [main] INFO org.apache.pig.t
2022-09-08 12:08:19,877 [main] INFO org.apache.pig.n
2022-09-08 12:08:19,892 [main] INFO org.apache.pig.b
2022-09-08 12:08:19,897 [main] INFO org.apache.pig.b
2022-09-08 12:08:19,897 [main] INFO org.apache.pig.b

2022-09-08 12:08:32,727 [main] INFO org.apache.pig.backend
(GOA,{(GOA,RAPE (SECTION 376 IPC),61)})
(ASSAM,{(ASSAM,RAPE (SECTION 376 IPC),1626)})
(BIHAR,{(BIHAR,RAPE (SECTION 376 IPC),1327)})
(DELHI,{(DELHI,RAPE (SECTION 376 IPC),892)})
(KERALA,{(KERALA,RAPE (SECTION 376 IPC),1259)})
(ODISHA,{(ODISHA,RAPE (SECTION 376 IPC),1666)})
(PUNJAB,{(PUNJAB,RAPE (SECTION 376 IPC),895)})
(SIKKIM,{(SIKKIM,RAPE (SECTION 376 IPC),29)})
(GUJARAT,{(GUJARAT,RAPE (SECTION 376 IPC),647)})
(HARYANA,{(HARYANA,RAPE (SECTION 376 IPC),940)})
(MANIPUR,{(MANIPUR,RAPE (SECTION 376 IPC),46)})
(MIZORAM,{(MIZORAM,RAPE (SECTION 376 IPC),122)})
(TRIPURA,{(TRIPURA,RAPE (SECTION 376 IPC),202)})
(NAGALAND,{(NAGALAND,RAPE (SECTION 376 IPC),26)})
(JHARKHAND,{(JHARKHAND,RAPE (SECTION 376 IPC),780)})
(KARNATAKA,{(KARNATAKA,RAPE (SECTION 376 IPC),842)})
(MEGHALAYA,{(MEGHALAYA,RAPE (SECTION 376 IPC),182)})
(RAJASTHAN,{(RAJASTHAN,RAPE (SECTION 376 IPC),1807)})
(CHANDIGARH,{(CHANDIGARH,RAPE (SECTION 376 IPC),34)})
(PUDUCHERRY,{(PUDUCHERRY,RAPE (SECTION 376 IPC),15)})
(TAMIL NADU,{(TAMIL NADU,RAPE (SECTION 376 IPC),962)})
(DAMAN & DIU,{(DAMAN & DIU,RAPE (SECTION 376 IPC),10)})

```

Output:

```

(UTTAR PRADESH,{(UTTAR PRADESH,RAPE (SECTION 376 IPC),3593)})
(ANDHRA PRADESH,{(ANDHRA PRADESH,RAPE (SECTION 376 IPC),1664)})
(MADHYA PRADESH,{(MADHYA PRADESH,RAPE (SECTION 376 IPC),4822)})
(JAMMU & KASHMIR,{(JAMMU & KASHMIR,RAPE (SECTION 376 IPC),388)})
(HIMACHAL PRADESH,{(HIMACHAL PRADESH,RAPE (SECTION 376 IPC),259)})

```


6. Calculate the number of males (age between 45-60 years) who were victims in “Kidnapping and abduction” in different states.

```
L1 = LOAD '/user/training/vaibhav/crimeage1.csv' using PigStorage(',') AS
(State, Crimehead, MB18, FB18, MB1830, FB1830, MB3035, FB3045, MB4560, FB4560, MA60, FA60,
TOTALMALE, TOTALFEMALE);

grunt> L1 = LOAD '/user/training/vaibhav/crimeage1.csv' using PigStorage(',')
AS (State, Crimehead, MB18, FB18, MB1830, FB1830, MB3035, FB3045, MB4560, FB4560, MA60, FA60, TOTALMALE, TOTALFEMALE);
2022-09-08 11:18:16,644 [main] WARN org.apache.hadoop.conf.Configuration - io.bytes.per.checksum is deprecated
2022-09-08 11:18:16,644 [main] WARN org.apache.hadoop.conf.Configuration - dfs.permissions.supergroup is deprecated
2022-09-08 11:18:16,644 [main] WARN org.apache.hadoop.conf.Configuration - dfs.max.objects is deprecated. Instead
2022-09-08 11:18:16,644 [main] WARN org.apache.hadoop.conf.Configuration - dfs.replication.interval is deprecated
2022-09-08 11:18:16,644 [main] WARN org.apache.hadoop.conf.Configuration - dfs.data.dir is deprecated. Instead

Y = FILTER L1 BY (Crimehead matches '("KIDNAPPING AND ABDUCTION").');
grunt> Y = FILTER L1 BY (Crimehead matches '("KIDNAPPING AND ABDUCTION").');
2022-09-08 11:18:27,195 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_CH
C = FOREACH Y generate Crimehead, MB4560; PA =
GROUP C ALL;
QA = FOREACH PA GENERATE SUM(C.$1);

grunt> C = FOREACH Y generate Crimehead, MB4560;
2022-09-08 11:19:04,329 [main] WARN org.apache.pig.PigServer
grunt> PA = GROUP C ALL;
2022-09-08 11:19:18,094 [main] WARN org.apache.pig.PigServer
grunt> QA = FOREACH PA GENERATE SUM(C.$1);
```

```
-----
(1022)
```

```
grunt>
```

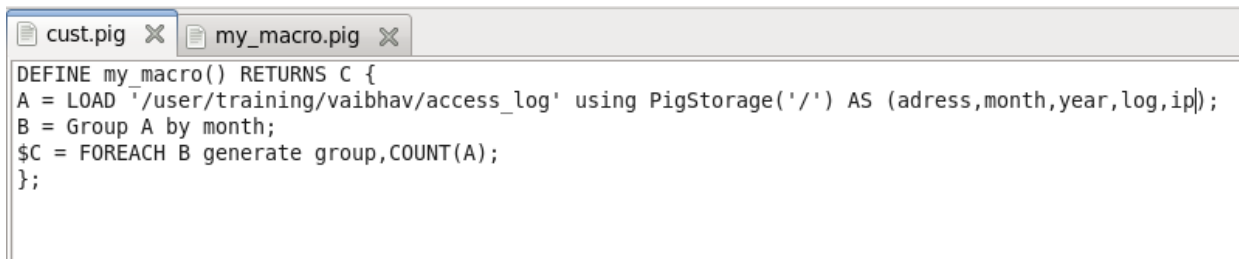
Assignment

Web log + Cust Analysis using Pig

Date:07-09-2022

Find Out the hit count per month

```
DEFINE my_macro() RETURNS C {  
A = LOAD '/user/training/vaibhav/access_log' using PigStorage('/') AS(adress,month,year,log,ip);  
B = Group A by month;  
$C = FOREACH B generate group,COUNT(A);  
};
```



The screenshot shows a code editor with two tabs: 'cust.pig' and 'my_macro.pig'. The 'my_macro.pig' tab is active, displaying the following Pig script:

```
DEFINE my_macro() RETURNS C {  
A = LOAD '/user/training/vaibhav/access_log' using PigStorage('/') AS (adress,month,year,log,ip);  
B = Group A by month;  
$C = FOREACH B generate group,COUNT(A);  
};
```

```
import '/desktop/cust.pig';
Count = my_macro();
```

```
grunt> import '/home/training/Desktop/cust.pig';
2022-09-07 10:01:15,370 [main] WARN org.apache.hadoop.conf.Configuration - dfs.
df.interval is deprecated. Instead, use fs.df.interval
2022-09-07 10:01:15,370 [main] WARN org.apache.hadoop.conf.Configuration - dfs.
max.objects is deprecated. Instead, use dfs.namenode.max.objects
2022-09-07 10:01:15,370 [main] WARN org.apache.hadoop.conf.Configuration - hadop.
native.lib is deprecated. Instead, use io.native.lib.available
2022-09-07 10:01:15,370 [main] WARN org.apache.hadoop.conf.Configuration - dfs.
data.dir is deprecated. Instead, use dfs.datanode.data.dir
2022-09-07 10:01:15,370 [main] WARN org.apache.hadoop.conf.Configuration - dfs.
name.dir is deprecated. Instead, use dfs.namenode.name.dir
2022-09-07 10:01:15,370 [main] WARN org.apache.hadoop.conf.Configuration - fs.d
efault.name is deprecated. Instead, use fs.defaultFS
2022-09-07 10:01:15,370 [main] WARN org.apache.hadoop.conf.Configuration - fs.c
heckpoint.dir is deprecated. Instead, use dfs.namenode.checkpoint.dir
2022-09-07 10:01:15,370 [main] WARN org.apache.hadoop.conf.Configuration - dfs.
block.size is deprecated. Instead, use dfs.blocksize
2022-09-07 10:01:15,370 [main] WARN org.apache.hadoop.conf.Configuration - dfs.
access.time.precision is deprecated. Instead, use dfs.namenode.accesstime.precis
ion
2022-09-07 10:01:15,370 [main] WARN org.apache.hadoop.conf.Configuration - dfs.
replication.min is deprecated. Instead, use dfs.namenode.replication.min
2022-09-07 10:01:15,370 [main] WARN org.apache.hadoop.conf.Configuration - dfs.
name.edits.dir is deprecated. Instead, use dfs.namenode.edits.dir
dump count;
```

```
grunt> count = my_macro();
grunt> dump count;
2022-09-07 10:02:53,950 [main] INFO org.apache.pig.tools.pigst
Pig features used in the script: GROUP_BY
2022-09-07 10:02:54,062 [main] INFO org.apache.pig.backend.had
ne.mapReduceLayer.MRCompiler - File concatenation threshold: 10
se
2022-09-07 10:02:54,072 [main] INFO org.apache.pig.backend.had
ne.mapReduceLayer.CombinerOptimizer - Choosing to move algebrai
ner
2022-09-07 10:02:54,126 [main] INFO org.apache.pig.backend.had
ne.mapReduceLayer.MultiQueryOptimizer - MR plan size before opt
```

Output:

```
2022-09-07 10:03:57,600 [main] INFO  org.apache.pig.backe
.MapReduceLauncher - Success!
2022-09-07 10:03:57,612 [main] INFO  org.apache.hadoop.ma
l input paths to process : 1
2022-09-07 10:03:57,612 [main] INFO  org.apache.pig.backe
l - Total input paths to process : 1
(Apr,301451)
(Aug,470275)
(Dec,186973)
(Feb,350885)
(Jan,273096)
(Jul,445653)
(Jun,386088)
(Mar,460257)
(May,339551)
(Nov,421626)
(Oct,411456)
(Sep,430529)
( [26,1)
(* [28,1)
(2|pq{jk@-1.-@lvo))1.(.1--'1()@+)* [29,1)
grunt> █
```

Find Out the hit count per status messages (2XX, 3XX, 4XX, 5XX)

Log= LOAD '/user/training/vaibhav/access_log' using PigStorage(' ')
as(ip:chararray, hp1:chararray, hp2:chararray, date:chararray, timezone:chararray, request1:chararray, request2:chararray, request3:chararray, code:int, bt:chararray);

B= GROUP Log by code;
CC = FOREACH B generate group, COUNT(Log);
Rcc = FILTER CC BY (\$0>400 AND \$0<500);
Pdm = GROUP Rcc ALL;
final = FOREACH Pdm GENERATE SUM(Rcc.\$1);

```
grunt> Log= LOAD '/user/training/vaibhav/access_log' using PigStorage(' ') as(ip:chararray, hp1:chararray, hp2:chararray, date:chararray, timezone:chararray, request1:chararray, request2:chararray, request3:chararray, code:int, bt:chararray);
2022-09-07 10:52:17,252 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_INT 2 time(s).
grunt> describe Log;
2022-09-07 10:52:23,809 [main] ERROR org.apache.pig.tools.grunt.Grunt - ERROR 1005: No plan for log to describe
Details at logfile: /home/training/pig_1662524156742.log
grunt> describe Log;
2022-09-07 10:52:30,529 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_INT 2 time(s).
Log: {ip: chararray, hp1: chararray, hp2: chararray, date: chararray, timezone: chararray, request1: chararray, request2: chararray, request3: chararray, code: int, bt: chararray}
grunt> B= GROUP Log by code;
2022-09-07 10:53:42,120 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_INT 2 time(s).
grunt> summ = FOREACH B generate group, COUNT( Log );
2022-09-07 10:53:56,194 [main] WARN org.apache.pig.PigServer - Encountered Warning IMPLICIT_CAST_TO_INT 2 time(s).
grunt> dump summ;
2022-09-07 10:54:04,474 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP_BY
grunt> B= GROUP Log by code;
grunt> CC = FOREACH B generate group, COUNT( Log );
grunt> Rcc = FILTER CC BY ($0>400 AND $0<500);
grunt> Pdm = GROUP Rcc ALL;
grunt>
-----
grunt> describe Pdm;
Pdm: {group: chararray, Rcc: {(group: int, long)}}
grunt> describe Rcc;
Rcc: {group: int, long}
grunt>
-----
grunt> final = FOREACH Pdm GENERATE SUM(Rcc.$1);
grunt> dump final;
2022-09-07 12:05:38,741 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pi
2022-09-07 12:05:38,893 [main] INFO org.apache.pig.backend.hadoop.executionengine.
2022-09-07 12:05:38,902 [main] INFO org.apache.pig.backend.hadoop.executionengine.
2022-09-07 12:05:38,908 [main] INFO org.apache.pig.backend.hadoop.executionengine.
2022-09-07 12:05:38,928 [main] INFO org.apache.pig.backend.hadoop.executionengine.
2022-09-07 12:05:38,928 [main] INFO org.apache.pig.backend.hadoop.executionengine.
2022-09-07 12:05:39,004 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pi
2022-09-07 12:05:39,010 [main] INFO org.apache.pig.backend.hadoop.executionengine.
2022-09-07 12:05:39,014 [main] INFO org.apache.pig.backend.hadoop.executionengine.
```

Output:

```
2022-09-07 12:06:55,046 [main] WARN  org.apache.pig.backend.hadoop
2022-09-07 12:06:55,046 [main] INFO  org.apache.pig.backend.hadoop
2022-09-07 12:06:55,050 [main] INFO  org.apache.hadoop.mapreduce
2022-09-07 12:06:55,050 [main] INFO  org.apache.pig.backend.hadoop
(84970)
grunt> █
```

Customer dataset:**Find out age group-wise customer count**

```

L = LOAD 'user/training/vaibhav/custs' using PigStorage(',') AS (id,fname,lname,age,profession);X =
FILTER L BY (age>30 AND age<=40);
G = GROUP X BY age;
C = FOREACH G generate group,COUNT(X);
Mapp = GROUP C ALL;
describe Mapp;
dump Mapp;
N = FOREACH Mapp GENERATE SUM(C.$1);
dump N;


```

```

grunt> L = LOAD 'user/training/vaibhav/custs' using PigStorage(',') AS (id,fname,lname,age,profession);
2022-09-07 11:33:54,114 [main] WARN org.apache.hadoop.conf.Configuration - io.bytes.per.checksum is deprecated. Inst
2022-09-07 11:33:54,114 [main] WARN org.apache.hadoop.conf.Configuration - dfs.permissions.supergroup is deprecated.
2022-09-07 11:33:54,114 [main] WARN org.apache.hadoop.conf.Configuration - dfs.max.objects is deprecated. Instead, u
2022-09-07 11:33:54,114 [main] WARN org.apache.hadoop.conf.Configuration - dfs.replication.interval is deprecated. I
2022-09-07 11:33:54,114 [main] WARN org.apache.hadoop.conf.Configuration - dfs.data.dir is deprecated. Instead, use
2022-09-07 11:33:54,114 [main] WARN org.apache.hadoop.conf.Configuration - dfs.access.time.precision is deprecated.
2022-09-07 11:33:54,114 [main] WARN org.apache.hadoop.conf.Configuration - dfs.replication.min is deprecated. Inst

```

```

grunt> X = FILTER L BY (age>30 AND age<=40);
2022-09-07 11:51:59,931 [main] WARN org.apache.
grunt> 

```

```

grunt> G = GROUP X BY age;
2022-09-07 11:52:25,426 [main] WARN
grunt> C = FOREACH G generate group,(
2022-09-07 11:52:30,733 [main] WARN
grunt> Mapp = GROUP C ALL;

```

```

2022-09-07 11:39:11,571 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.Ma
2022-09-07 11:39:11,575 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total i
2022-09-07 11:39:11,575 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil -
(all,{(21,114),(22,100),(23,108),(24,110),(25,176),(26,210),(27,181),(28,190),(29,203),(30,211)})
grunt> █

```

```

grunt> N = FOREACH Mapp GENERATE SUM(C.$1);
2022-09-07 11:37:36,756 [main] WARN org.apache.pig.
grunt> █

```

```

grunt> dump N;
2022-09-07 11:37:58,423 [main] WARN org.apache.pig.PigServer - Enc
2022-09-07 11:37:58,424 [main] INFO org.apache.pig.tools.pigstats.
2022-09-07 11:37:58,551 [main] INFO org.apache.pig.backend.hadoop.
2022-09-07 11:37:58,562 [main] INFO org.apache.pig.backend.hadoop.
2022-09-07 11:37:58,564 [main] INFO org.apache.pig.backend.hadoop.
2022-09-07 11:37:58,580 [main] INFO org.apache.pig.backend.hadoop.
2022-09-07 11:37:58,580 [main] INFO org.apache.pig.backend.hadoop.
2022-09-07 11:37:58,681 [main] INFO org.apache.pig.tools.pigstats.
2022-09-07 11:37:58,685 [main] INFO org.apache.pig.backend.hadoop.
2022-09-07 11:37:58,686 [main] INFO org.apache.pig.backend.hadoop.
2022-09-07 11:38:00,946 [main] INFO org.apache.pig.backend.hadoop.
2022-09-07 11:38:00,957 [main] INFO org.apache.pig.backend.hadoop.
2022-09-07 11:38:00,977 [main] INFO org.apache.pig.backend.hadoop.
2022-09-07 11:38:00,977 [main] INFO org.apache.pig.backend.hadoop.
f reducers to 1

```

Output:

```

2022-09-07 11:53:44,856 [main] INFO
2022-09-07 11:53:44,858 [main] INFO
2022-09-07 11:53:44,858 [main] INFO
(2061)
grunt> █

```