

INDEX

Sr. No	Assignment Title	Date	Page No
1	Write a program to find occurrence of any specific word	08-09-2022	2-6
2	Write a program to find average word length.	10-08-2022	7-10
3	Encrypt the email ids	23-08-2022	11-13
4	Multiple file input demo assignment	29-08-2022	14-18
5	Web Log Analysis - partitioner demo assignment I. Odd roll no II. Even roll No	30-08-2022	19-33
6	Partitioner Demo - Cust dataset	02-09-2022	34-40
7	Government Crime Assignment - Pig	08-09-2022	41-45

Assignment : 01

Write a program to find occurrence of any specific word

Code:

```
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.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 Count {
    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();
            String find = "big";
            for (String word : line.split("\\W+")){
                if(find.equals(word.toLowerCase())){
                    context.write(new Text(find), 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));
        }
    }
}
```

```

    }
}

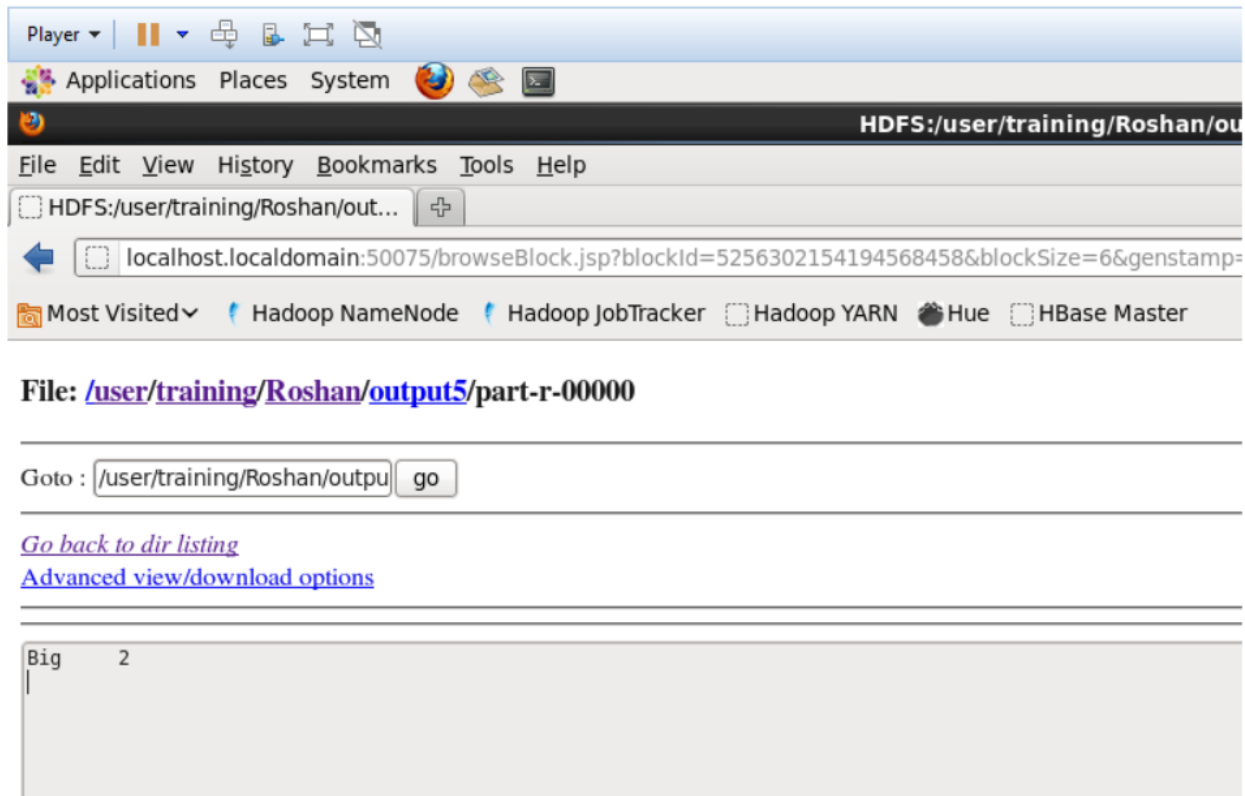
public static void main(String[] args) throws Exception{
    if(args.length != 2){
        System.out.printf("Usage: WordCount <input dir> <output dir>\n");
        System.exit(-1);
    }
    Job job = new Job();
    job.setJarByClass(Count.class);
    job.setJobName("WordCount");
    FileInputFormat.setInputPaths(job,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

```

```

[training@localhost ~]$ hadoop jar /home/training/Custs8.jar /user/training/custs /user/training/Roshan/cust-op8
22/08/08 11:08:08 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
22/08/08 11:08:08 INFO input.FileInputFormat: Total input paths to process : 1
22/08/08 11:08:08 WARN snappy.LoadSnappy: Snappy native library is available
22/08/08 11:08:08 INFO snappy.LoadSnappy: Snappy native library loaded
22/08/08 11:08:09 INFO mapred.JobClient: Running job: job_202208080931_0014
22/08/08 11:08:10 INFO mapred.JobClient: map 0% reduce 0%
22/08/08 11:08:13 INFO mapred.JobClient: map 100% reduce 0%
22/08/08 11:08:16 INFO mapred.JobClient: map 100% reduce 100%
22/08/08 11:08:16 INFO mapred.JobClient: Job complete: job_202208080931_0014
22/08/08 11:08:16 INFO mapred.JobClient: Counters: 32
22/08/08 11:08:16 INFO mapred.JobClient:   File System Counters
22/08/08 11:08:16 INFO mapred.JobClient:     FILE: Number of bytes read=2514
22/08/08 11:08:16 INFO mapred.JobClient:     FILE: Number of bytes written=366752
22/08/08 11:08:16 INFO mapred.JobClient:     FILE: Number of read operations=0
22/08/08 11:08:16 INFO mapred.JobClient:     FILE: Number of large read operations=0
22/08/08 11:08:16 INFO mapred.JobClient:     FILE: Number of write operations=0
22/08/08 11:08:16 INFO mapred.JobClient:     HDFS: Number of bytes read=391459
22/08/08 11:08:16 INFO mapred.JobClient:     HDFS: Number of bytes written=10
22/08/08 11:08:16 INFO mapred.JobClient:     HDFS: Number of read operations=2
22/08/08 11:08:16 INFO mapred.JobClient:     HDFS: Number of large read operations=0
22/08/08 11:08:16 INFO mapred.JobClient:     HDFS: Number of write operations=1
22/08/08 11:08:16 INFO mapred.JobClient:   Job Counters
22/08/08 11:08:16 INFO mapred.JobClient:     Launched map tasks=1
22/08/08 11:08:16 INFO mapred.JobClient:     Launched reduce tasks=1
22/08/08 11:08:16 INFO mapred.JobClient:     Data-local map tasks=1
22/08/08 11:08:16 INFO mapred.JobClient:     Total time spent by all maps in occupied slots (ms)=2974
22/08/08 11:08:16 INFO mapred.JobClient:     Total time spent by all reduces in occupied slots (ms)=2273
22/08/08 11:08:16 INFO mapred.JobClient:     Total time spent by all maps waiting after reserving slots (ms)=0

```



Count of specific profession

Code:

```
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.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 Cust {
    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();
            String find = "Pilot";
```

```

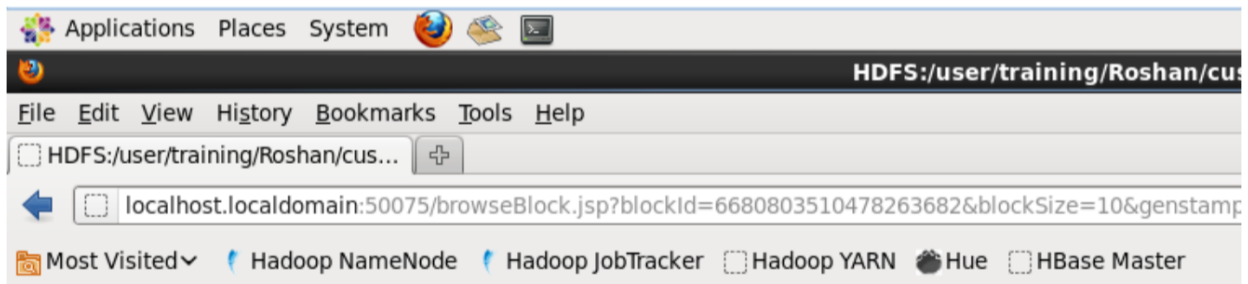
        for (String word : line.split(" ")) {
            if (find.equals(word)) {
                context.write(new Text(find), new IntWritable(1));
            }
        }
    }
}

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: WordCount <input dir> <output dir>\n");
        System.exit(-1);
    }
    Job job = new Job();
    job.setJarByClass(Cust.class);
    job.setJobName("Custs");
    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);
}
}

```

```
[training@localhost ~]$ hadoop jar /home/training/Custs8.jar /user/training/custs /user/training/Roshan/cust-op8
22/08/08 11:08:08 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
22/08/08 11:08:08 INFO input.FileInputFormat: Total input paths to process : 1
22/08/08 11:08:08 WARN snappy.LoadSnappy: Snappy native library is available
22/08/08 11:08:08 INFO snappy.LoadSnappy: Snappy native library loaded
22/08/08 11:08:09 INFO mapred.JobClient: Running job: job_202208080931_0014
22/08/08 11:08:10 INFO mapred.JobClient: map 0% reduce 0%
22/08/08 11:08:13 INFO mapred.JobClient: map 100% reduce 0%
22/08/08 11:08:16 INFO mapred.JobClient: map 100% reduce 100%
22/08/08 11:08:16 INFO mapred.JobClient: Job complete: job_202208080931_0014
22/08/08 11:08:16 INFO mapred.JobClient: Counters: 32
22/08/08 11:08:16 INFO mapred.JobClient:   File System Counters
22/08/08 11:08:16 INFO mapred.JobClient:     FILE: Number of bytes read=2514
22/08/08 11:08:16 INFO mapred.JobClient:     FILE: Number of bytes written=366752
22/08/08 11:08:16 INFO mapred.JobClient:     FILE: Number of read operations=0
22/08/08 11:08:16 INFO mapred.JobClient:     FILE: Number of large read operations=0
22/08/08 11:08:16 INFO mapred.JobClient:     FILE: Number of write operations=0
22/08/08 11:08:16 INFO mapred.JobClient:     HDFS: Number of bytes read=391459
22/08/08 11:08:16 INFO mapred.JobClient:     HDFS: Number of bytes written=10
22/08/08 11:08:16 INFO mapred.JobClient:     HDFS: Number of read operations=2
22/08/08 11:08:16 INFO mapred.JobClient:     HDFS: Number of large read operations=0
22/08/08 11:08:16 INFO mapred.JobClient:     HDFS: Number of write operations=1
22/08/08 11:08:16 INFO mapred.JobClient:   Job Counters
22/08/08 11:08:16 INFO mapred.JobClient:     Launched map tasks=1
22/08/08 11:08:16 INFO mapred.JobClient:     Launched reduce tasks=1
22/08/08 11:08:16 INFO mapred.JobClient:     Data-local map tasks=1
```



File: [/user/training/Roshan/cust-op8/part-r-00000](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

Pilot 209

Assignment : 02

Write a program to find average word length.

Code:

```
import java.io.IOException;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.FloatWritable;
import org.apache.hadoop.io.Text;
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, FloatWritable>
    {
        @Override
        public void map(LongWritable key, Text value, Context context)
        throws IOException, InterruptedException {
            Text firstLetter = new Text();
            FloatWritable wordLength = new FloatWritable();

            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,FloatWritable,Text,FloatWritable>{
public void reduce(Text key,Iterable<FloatWritable>values,Context context)
throws IOException,InterruptedException{
    int sum =0;
    int count =0;
    int Average =0;
    for(FloatWritable val:values) {
        sum += val.get();
        count = count+1;
    }
    Average = sum/count;

    context.write(key,new FloatWritable(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]));
}

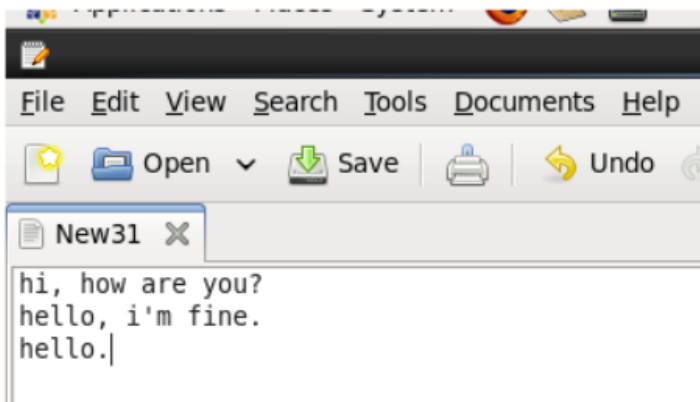
```



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

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

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





Average1.jar



Average2.jar

```
[training@localhost ~]$ hadoop jar /home/training/Average2.jar /user/training/Ro
shan/Avg /user/training/Roshan/output-avg2
22/08/17 10:20:09 WARN mapred.JobClient: Use GenericOptionsParser for parsing th
e arguments. Applications should implement Tool for the same.
22/08/17 10:20:09 INFO input.FileInputFormat: Total input paths to process : 1
22/08/17 10:20:09 WARN snappy.LoadSnappy: Snappy native library is available
22/08/17 10:20:09 INFO snappy.LoadSnappy: Snappy native library loaded
22/08/17 10:20:09 INFO mapred.JobClient: Running job: job_202208170950_0010
22/08/17 10:20:10 INFO mapred.JobClient: map 0% reduce 0%
22/08/17 10:20:13 INFO mapred.JobClient: map 100% reduce 0%
22/08/17 10:20:15 INFO mapred.JobClient: Job complete: job_202208170950_0010
22/08/17 10:20:15 INFO mapred.JobClient: Counters: 32
22/08/17 10:20:15 INFO mapred.JobClient:   File System Counters
22/08/17 10:20:15 INFO mapred.JobClient:       FILE: Number of bytes read=70
22/08/17 10:20:15 INFO mapred.JobClient:       FILE: Number of bytes written=36190
6
22/08/17 10:20:15 INFO mapred.JobClient:       FILE: Number of read operations=0
22/08/17 10:20:15 INFO mapred.JobClient:       FILE: Number of large read operatio
ns=0
22/08/17 10:20:15 INFO mapred.JobClient:       FILE: Number of write operations=0
22/08/17 10:20:15 INFO mapred.JobClient:       HDFS: Number of bytes read=150
22/08/17 10:20:15 INFO mapred.JobClient:       HDFS: Number of bytes written=30
22/08/17 10:20:15 INFO mapred.JobClient:       HDFS: Number of read operations=2
22/08/17 10:20:15 INFO mapred.JobClient:       HDFS: Number of large read operatio
ns=0
22/08/17 10:20:15 INFO mapred.JobClient:       HDFS: Number of write operations=1
22/08/17 10:20:15 INFO mapred.JobClient: Job Counters
```

File Edit View History Bookmarks Tools Help

☐ HDFS:/user/training/Roshan/out...

☐ localhost.localdomain:50075/browseBlock.jsp?blockId=-6230

Most Visited Hadoop NameNode Hadoop JobTracker ☐ H...

Goto :

[Go back to dir listing](#)
[Advanced view/download options](#)

a	3.0
f	5.0
h	4.0
i	3.0
y	4.0

Assignment : 03

Encrypt the email ids

use "Only mapper" design pattern

```
import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.util.GenericOptionsParser;

public class Encry {

    public static class MapForEncryption extends Mapper<LongWritable, Text, Text, Text>
    {
        public void map(LongWritable key, Text Value, Context con)throws IOException,
        InterruptedException
        {
            String line=Value.toString();

            int encry;           //created a object which will store he encrypted charater
            String mask = ""; //mask object initialized with null value which will store the actual output text
            for(int i=0;i<line.length();i++)    //for loop which will iterate through line
            {
                char c = line.charAt(i);    //store each character at index i

                encry = 4 + (int)c;        //increment the alphabet by 4

                c = (char)encry; //type cast encry into char
                mask += c;      //store the final text in mask
            }
            //String line1 = line.replace("||", ",");
            Text outputKey = new Text(mask);
            Text outputValue = new Text("");
            con.write(outputKey,outputValue);
        }
    }

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

```

Configuration c = new Configuration();
String[] files = new GenericOptionsParser(c,args).getRemainingArgs();
Path input = new Path(files[0]);
Path output = new Path(files[1]);

Job j = new Job(c,"FormatEmpFile");

j.setJarByClass(Encry.class);

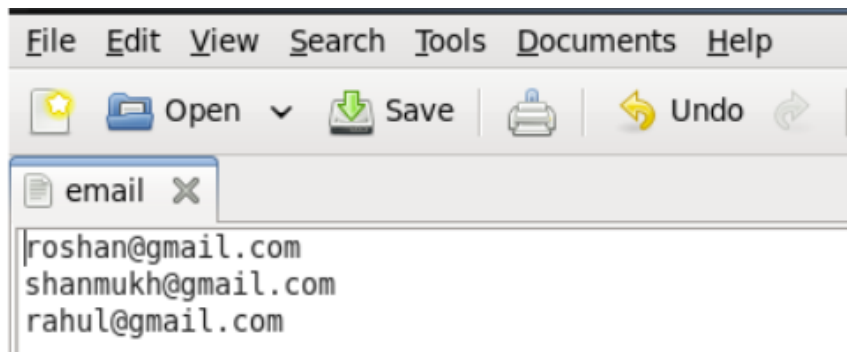
j.setJobName("");

j.setMapperClass(MapForEncryption.class);
j.setNumReduceTasks(0);

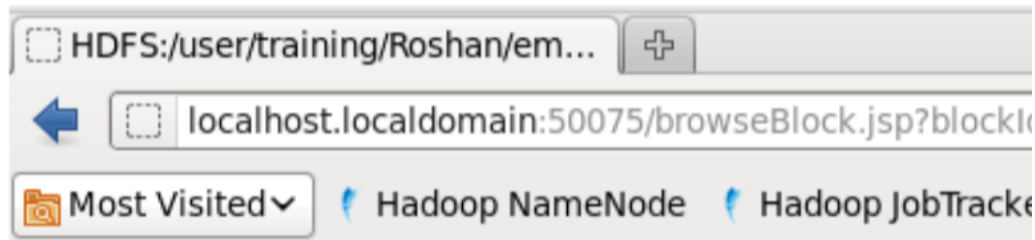
j.setMapOutputKeyClass(Text.class);
j.setMapOutputValueClass(Text.class);

FileInputFormat.addInputPath(j,input);
FileOutputFormat.setOutputPath(j,output);
System.exit(j.waitForCompletion(true)? 0:1);
}}

```



```
[training@localhost ~]$ hadoop jar /home/training/encry.jar /user/training/Roshan/email /user/training/Roshan/email-op9
22/08/23 11:51:49 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement
22/08/23 11:51:49 INFO input.FileInputFormat: Total input paths to process : 1
22/08/23 11:51:49 WARN snappy.LoadSnappy: Snappy native library is available
22/08/23 11:51:49 INFO snappy.LoadSnappy: Snappy native library loaded
22/08/23 11:51:50 INFO mapred.JobClient: Running job: job_202208230930_0016
22/08/23 11:51:51 INFO mapred.JobClient: map 0% reduce 0%
22/08/23 11:51:55 INFO mapred.JobClient: map 100% reduce 0%
22/08/23 11:51:56 INFO mapred.JobClient: Job complete: job_202208230930_0016
22/08/23 11:51:56 INFO mapred.JobClient: Counters: 24
22/08/23 11:51:56 INFO mapred.JobClient:   File System Counters
22/08/23 11:51:56 INFO mapred.JobClient:     FILE: Number of bytes read=0
22/08/23 11:51:56 INFO mapred.JobClient:     FILE: Number of bytes written=180563
22/08/23 11:51:56 INFO mapred.JobClient:     FILE: Number of read operations=0
22/08/23 11:51:56 INFO mapred.JobClient:     FILE: Number of large read operations=0
22/08/23 11:51:56 INFO mapred.JobClient:     FILE: Number of write operations=0
22/08/23 11:51:56 INFO mapred.JobClient:     HDFS: Number of bytes read=163
22/08/23 11:51:56 INFO mapred.JobClient:     HDFS: Number of bytes written=55
22/08/23 11:51:56 INFO mapred.JobClient:     HDFS: Number of read operations=2
22/08/23 11:51:56 INFO mapred.JobClient:     HDFS: Number of large read operations=0
22/08/23 11:51:56 INFO mapred.JobClient:     HDFS: Number of write operations=1
```



File: [/user/training/Roshan/email-op9/part-m-00000](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

```
vswlerDkqemp2gsq
wlerqyolDkqemp2gsq
velypDkqemp2gsq
```

Assignment : 04

Multiple file input demo assignment

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

```
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 MultipleInputFiles {

    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[2]);
        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[1];
        Text outputKey = new Text(name);
        int cost = Integer.parseInt(line1[3]);
        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

    {
//to get the sum of cost
//          int sum = 0;

```

```

//                for(IntWritable value : total_cost){
//                    sum += value.get();
//                }

//to get the max cost
                int max = 0;

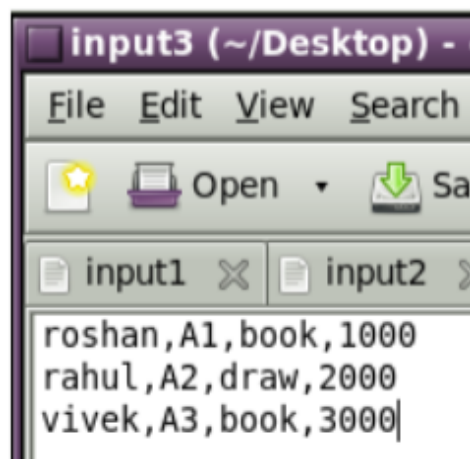
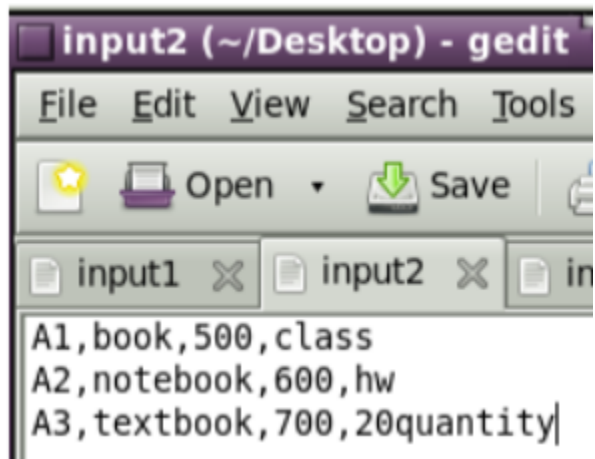
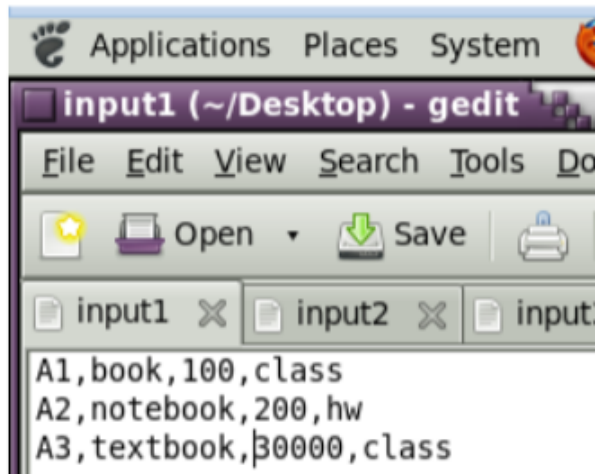
                for (IntWritable value : total_cost) {
                    if (value.get() > max) {
                        max = value.get();
                    }
                }
                con.write(name,new IntWritable(max));
            }
        }

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(MultipleInputFiles.class);
    j.setMapperClass(map1.class);
    j.setMapperClass(Map2.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);
}

```

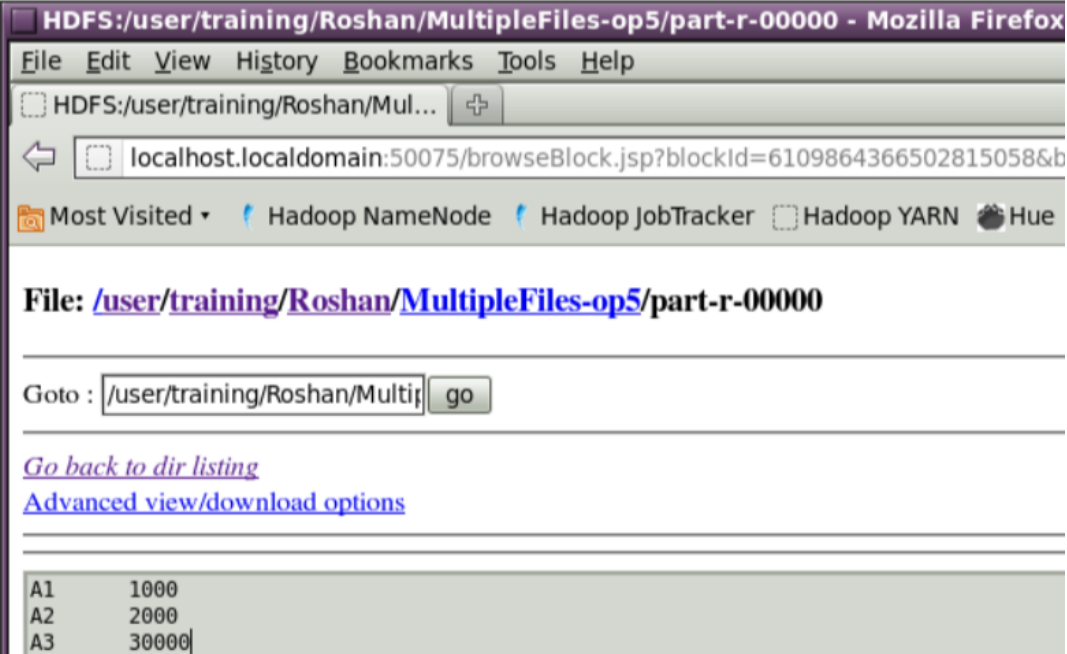


```
}  
  
}
```



```
[training@localhost ~]$ hadoop jar /home/training/MultipleInput.jar /user/traini
ng/Roshan/Multiple-ipp1 /user/training/Roshan/Multiple-ip2 /user/training/Roshan
/Multiple-ip3 /user/training/Roshan/MultipleFiles-op5
22/08/29 10:47:49 WARN mapred.JobClient: Use GenericOptionsParser for parsing th
e arguments. Applications should implement Tool for the same.
22/08/29 10:47:49 INFO input.FileInputFormat: Total input paths to process : 1
22/08/29 10:47:49 WARN snappy.LoadSnappy: Snappy native library is available
22/08/29 10:47:49 INFO snappy.LoadSnappy: Snappy native library loaded
22/08/29 10:47:49 INFO input.FileInputFormat: Total input paths to process : 1
22/08/29 10:47:49 INFO input.FileInputFormat: Total input paths to process : 1
22/08/29 10:47:49 INFO mapred.JobClient: Running job: job_202208290924_0005
22/08/29 10:47:50 INFO mapred.JobClient: map 0% reduce 0%
22/08/29 10:47:54 INFO mapred.JobClient: map 66% reduce 0%
22/08/29 10:47:57 INFO mapred.JobClient: map 100% reduce 0%
22/08/29 10:47:58 INFO mapred.JobClient: map 100% reduce 100%
22/08/29 10:47:59 INFO mapred.JobClient: Job complete: job_202208290924_0005
22/08/29 10:47:59 INFO mapred.JobClient: Counters: 32
22/08/29 10:47:59 INFO mapred.JobClient: File System Counters
22/08/29 10:48:00 INFO mapred.JobClient: FILE: Number of bytes read=87
22/08/29 10:48:00 INFO mapred.JobClient: FILE: Number of bytes written=73445
4
22/08/29 10:48:00 INFO mapred.JobClient: FILE: Number of read operations=0
22/08/29 10:48:00 INFO mapred.JobClient: FILE: Number of large read operatio
ns=0
```

For max cost:



HDFS:/user/training/Roshan/MultipleFiles-op5/part-r-00000 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

HDFS:/user/training/Roshan/Mul... +

localhost.localdomain:50075/browseBlock.jsp?blockId=6109864366502815058&b

Most Visited Hadoop NameNode Hadoop JobTracker Hadoop YARN Hue

File: /user/training/Roshan/MultipleFiles-op5/part-r-00000

Goto : /user/training/Roshan/Multi go

[Go back to dir listing](#)

[Advanced view/download options](#)

A1	1000
A2	2000
A3	30000

Assignment : 05

Web Log Analysis - partitioner demo assignment

Odd

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 part 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[] date =line.split("/");
            String month=date[1];
            context.write(new Text (month), new IntWritable (1));
        }
    }

    //Reducer class
    public static class SumReducer extends
Reducer<Text,IntWritable,Text,IntWritable> {
        @Override public void reduce(Text name, Iterable<IntWritable>values,Context
context)
        throws IOException, InterruptedException {
            int Count = 0;
            for (IntWritable value : values) {
                Count += value.get();
            }
        }
    }
}
```

```

context.write(name, new IntWritable (Count));
}
}

```

//Partitioner class

```

public static class CaderPartitioner extends Partitioner<Text,IntWritable>{
public int getPartition (Text name,IntWritable value,int numReduceTasks)
{

    if (name.toString().equals("Jan"))
    {
        return 1 ;
    }
    else if (name.toString().equals("Feb"))
    {
        return 2;
    }
    else if (name.toString().equals("Mar") )
    {
        return 3;
    }
    else if (name.toString().equals("Apr") )
    {
        return 4;    }
    else if (name.toString().equals("May") )
    {
        return 5;
    }
    else if (name.toString().equals("Jun") )
    {
        return 6;
    }
    else if (name.toString().equals("Jul") )
    {
        return 7;
    }
    else if (name.toString().equals("Aug"))
    {
        return 8;
    }
    else if (name.toString().equals("Sep"))
    {
        return 9;
    }
    else if (name.toString().equals("Oct"))

```

```

    {
        return 10;
    }
    else if (name.toString().equals("Nov"))
    {
        return 11;
    }
    else if (name.toString().equals("Dec"))
    {
        return 12;
    }
    else{
        return 0;
    }
}

}

}

public static void main(String[] arg) throws Exception{
// TODO Auto-generated method stub
int res=ToolRunner.run(new Configuration(),new part(), arg);
System.exit(0);
}

@Override
public int run(String[] arg) throws Exception {
// TODO Auto-generated method stub
Configuration conf = getConf() ;
Job job = new Job(conf,"part");
job.setJarByClass(part.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 ;
}
}

```

}

HDFS:/user/training/roshan31/lo... +

localhost.localdomain:50075/browseBlock.jsp?blockId=2483212335204478153&blockSize=67108864&genstamp=2407&file

Most Visited • Hadoop NameNode Hadoop JobTracker Hadoop YARN Hue HBase Master

File: /user/training/roshan31/logs

Goto : /user/training/roshan31 go

[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/closetlabel.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 -
```

```
[training@localhost ~]$ hadoop jar /home/training/p1.jar /user/training/roshan31/logs /user/training/roshan31/part/partitioner031
22/09/01 11:58:44 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
22/09/01 11:58:44 INFO input.FileInputFormat: Total input paths to process : 1
22/09/01 11:58:44 WARN snappy.LoadSnappy: Snappy native library is available
22/09/01 11:58:44 INFO snappy.LoadSnappy: Snappy native library loaded
22/09/01 11:58:44 INFO mapred.JobClient: Running job: job_202209010917_0034
22/09/01 11:58:45 INFO mapred.JobClient: map 0% reduce 0%
22/09/01 11:58:56 INFO mapred.JobClient: map 19% reduce 0%
22/09/01 11:58:57 INFO mapred.JobClient: map 21% reduce 0%
22/09/01 11:58:59 INFO mapred.JobClient: map 25% reduce 0%
22/09/01 11:59:08 INFO mapred.JobClient: map 35% reduce 1%
22/09/01 11:59:09 INFO mapred.JobClient: map 50% reduce 1%
22/09/01 11:59:11 INFO mapred.JobClient: map 50% reduce 2%
22/09/01 11:59:17 INFO mapred.JobClient: map 71% reduce 2%
22/09/01 11:59:19 INFO mapred.JobClient: map 72% reduce 2%
22/09/01 11:59:20 INFO mapred.JobClient: map 75% reduce 3%
22/09/01 11:59:27 INFO mapred.JobClient: map 98% reduce 3%
22/09/01 11:59:28 INFO mapred.JobClient: map 100% reduce 3%
22/09/01 11:59:29 INFO mapred.JobClient: map 100% reduce 4%
22/09/01 11:59:30 INFO mapred.JobClient: map 100% reduce 9%
22/09/01 11:59:31 INFO mapred.JobClient: map 100% reduce 15%
22/09/01 11:59:34 INFO mapred.JobClient: map 100% reduce 23%
```

Applications Places System

HDFS:/user/training/roshan31/part/partitioner031 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

HDFS:/user/training/roshan31/p...

localhost.localdomain:50075/browseDirectory.jsp?dir=%2Fuser%2Ftraining%2Froshan31%2Fpart%2Fpartitioner031&name

Most Visited Hadoop NameNode Hadoop JobTracker Hadoop YARN Hue HBase Master

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 11:59	rw-r--r--	training	supergroup
logs	dir				2022-09-01 11:58	rw-r--r--	training	supergroup
part-r-00000	file	0.05 KB	1	64 MB	2022-09-01 11:59	rw-r--r--	training	supergroup
part-r-00001	file	0.01 KB	1	64 MB	2022-09-01 11:59	rw-r--r--	training	supergroup
part-r-00002	file	0.01 KB	1	64 MB	2022-09-01 11:59	rw-r--r--	training	supergroup
part-r-00003	file	0.01 KB	1	64 MB	2022-09-01 11:59	rw-r--r--	training	supergroup
part-r-00004	file	0.01 KB	1	64 MB	2022-09-01 11:59	rw-r--r--	training	supergroup
part-r-00005	file	0.01 KB	1	64 MB	2022-09-01 11:59	rw-r--r--	training	supergroup
part-r-00006	file	0.01 KB	1	64 MB	2022-09-01 11:59	rw-r--r--	training	supergroup
part-r-00007	file	0.01 KB	1	64 MB	2022-09-01 11:59	rw-r--r--	training	supergroup
part-r-00008	file	0.01 KB	1	64 MB	2022-09-01 11:59	rw-r--r--	training	supergroup
part-r-00009	file	0.01 KB	1	64 MB	2022-09-01 11:59	rw-r--r--	training	supergroup
part-r-00010	file	0.01 KB	1	64 MB	2022-09-01 11:59	rw-r--r--	training	supergroup
part-r-00011	file	0.01 KB	1	64 MB	2022-09-01 11:59	rw-r--r--	training	supergroup
part-r-00012	file	0.01 KB	1	64 MB	2022-09-01 11:59	rw-r--r--	training	supergroup

[Go back to DFS home](#)

HDFS:/user/training/roshan31/part/partitioner031/part-r-00001 - Mozilla

File Edit View History Bookmarks Tools Help

HDFS:/user/training/roshan31/p... +

localhost.localdomain:50075/browseBlock.jsp?blockId=43632885786924784

Most Visited ▾ Hadoop NameNode Hadoop JobTracker Hadoop YARN

File: [/user/training/roshan31/part/partitioner031/part-r-00001](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

Jan	273096
-----	--------

File: [/user/training/roshan31/part/partitioner031/part-r-00002](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

Feb	350885
-----	--------

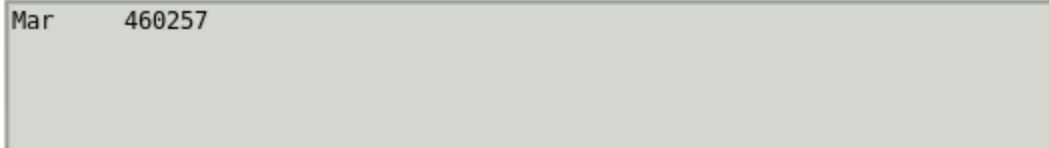
File: [/user/training/roshan31/part/partitioner031/part-r-00003](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

Mar 460257



File: [/user/training/roshan31/part/partitioner031/part-r-00004](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

Apr 301451



File: [/user/training/roshan31/part/partitioner031/part-r-00005](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

May 339551



File: [/user/training/roshan31/part/partitioner031/part-r-00006](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

Jun 386088

File: [/user/training/roshan31/part/partitioner031/part-r-00007](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

Jul 445653

File: [/user/training/roshan31/part/partitioner031/part-r-00008](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

Aug 470275

File: [/user/training/roshan31/part/partitioner031/part-r-00009](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

Sep 430529

File: [/user/training/roshan31/part/partitioner031/part-r-00010](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

Oct 411456

File: [/user/training/roshan31/part/partitioner031/part-r-00011](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

Nov 421626

File: [/user/training/roshan31/part/partitioner031/part-r-00012](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

Dec 186973

Even

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.Mapper.Context;
import org.apache.hadoop.mapreduce.lib.input.*;
import org.apache.hadoop.mapreduce.lib.output.*;
import org.apache.hadoop.util.*;

public class age extends Configured implements Tool {
    public static class Map extends Mapper<LongWritable, Text, Text, Text> {
        @Override public void map (LongWritable key, Text value, Context context)
            throws IOException, InterruptedException {
            String line = value.toString();
            String[] cust =line.split(",");
            String age=cust[3];
            context.write(new Text (age), new Text(value));
        }
    }

    public static class CodeReducer extends Reducer<Text, Text, Text, IntWritable>{
        @Override
        public void reduce(Text name, Iterable<Text> values, Context context) throws
        IOException, InterruptedException {
            int Count = 0;
            for (Text value: values){
                Count += 1;
            }
            context.write(name, new IntWritable(Count));
        }
    }

    public static class CaderPartitioner extends Partitioner<Text,Text>{
        public int getPartition (Text name,Text value,int numReduceTasks)
        {
            String line = value.toString();
            String[] cust =line.split(",");
            int age=Integer.parseInt(cust[3]);

            if (age>=20 && age<=30)
```

```

{
    return 1 ;
}
else if (age>=31 && age<=45)
{
    return 2;
}
else if (age>=46 && age<=60 )
{
    return 3;
}
else if (age>=61 )
{
    return 4;
}

else
{
return 0;
}

}
}

public int run(String[] arg) throws Exception {
// TODO Auto-generated method stub
Configuration conf = getConf() ;
Job job = new Job(conf,"cust");
job.setJarByClass(age.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(Text.class) ;

job.setPartitionerClass(CaderPartitioner.class) ;
job.setReducerClass(CodeReducer.class);
job.setNumReduceTasks(6) ;
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 ;
    }
    public static void main(String[] arg) throws Exception{
        // TODO Auto-generated method stub
        int res=ToolRunner.run(new Configuration(),new age(), arg);
        System.exit(0);
    }
}
```

HDFS:/user/training/roshan31/c...

localhost.localdomain:50075/browseBlock.jsp?blockId=2992766822119786084&blockSize=70&genstamp=3718&fi

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

File: [/user/training/roshan31/cust-part/cust0310/part-r-00001](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

21	114
22	100
23	108
24	110
25	176
26	210
27	181
28	190
29	203
30	211

File: [/user/training/roshan31/cust-part/cust0310/part-r-00002](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

31	191
32	205
33	191
34	211
35	225
36	205
37	195
38	232
39	216
40	190
41	193
42	172
43	216
44	204
45	196

File: [/user/training/roshan31/cust-part/cust0310/part-r-00003](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

46	97
47	97
48	115
49	91
50	96
51	219
52	192
53	192
54	195
55	184
56	188
57	203
58	220
59	228
60	197

File: [/user/training/roshan31/cust-part/cust0310/part-r-00004](#)

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

61	181
62	193
63	211
64	223
65	213
66	201
67	202
68	173
69	181
70	194
71	194
72	196
73	171
74	213
75	94

Assignment : 06

Partitioner Demo - Cust dataset

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.Mapper.Context;
import org.apache.hadoop.mapreduce.lib.input.*;
import org.apache.hadoop.mapreduce.lib.output.*;
import org.apache.hadoop.util.*;

public class cust extends Configured implements Tool{
    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[] cust =line.split(",",-3);
            String name=cust[4];
            context.write(new Text (name), new IntWritable (1));
        }
    }

    public static class CodeReducer extends Reducer<Text, IntWritable, Text, IntWritable>{
        @Override
        public void reduce(Text name, Iterable<IntWritable> values, Context context)
            throws IOException, InterruptedException {
            int Count = 0;
            for (IntWritable value: values){
                Count += value.get();
            }
            context.write(name, new IntWritable(Count));
        }
    }

    public static class CaderPartitioner extends Partitioner<Text,IntWritable>{
```

```

public int getPartition (Text name,IntWritable value,int numReduceTasks)
{

    if (name.toString().equals("Pilot"))
    {
        return 1 ;
    }
    else if (name.toString().equals("Teacher"))
    {
        return 2;
    }
    else if (name.toString().equals("Lawyer" )
    {
        return 3;
    }
    else if (name.toString().equals("Pharmacist" )
    {
        return 4;    }
    else if (name.toString().equals("Carpenter" )
    {
        return 5;
    }

    else{
return 0;
    }

    }
}

```

```

public static void main(String[] arg) throws Exception{
    // TODO Auto-generated method stub
    int res=ToolRunner.run(new Configuration(),new cust(), arg);
    System.exit(0);
}

```

```

public int run(String[] arg) throws Exception {
    // TODO Auto-generated method stub
    Configuration conf = getConf() ;
    Job job = new Job(conf,"cust");
    job.setJarByClass(cust.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(CodeReducer.class);
job.setNumReduceTasks(6) ;
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 ;
}
}

```

HDFS:/user/training/custs

localhost.localdomain:50075/browseBlock.jsp?blockId=8965518627038833789&blockSize=3913556

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

File: /user/training/custs

Goto :

[Go back to dir listing](#)

[Advanced view/download options](#)

[View Next chunk](#)

```

4000001,Kristina,Chung,55,Pilot|
4000002,Paige,Chen,74,Teacher
4000003,Sherri,Melton,34,Firefighter
4000004,Gretchen,Hill,66,Computer hardware engineer
4000005,Karen,Puckett,74,Lawyer
4000006,Patrick,Song,42,Veterinarian
4000007,Elsie,Hamilton,43,Pilot
4000008,Hazel,Bender,63,Carpenter
4000009,Malcolm,Wagner,39,Artist
4000010,Dolores,McLaughlin,60,Writer
4000011,Francis,McNamara,47,Therapist
4000012,Sandy,Raynor,26,Writer
4000013,Marion,Moon,41,Carpenter
4000014,Beth,Woodard,65,
4000015,Julia,Desai,49,Musician
4000016,Jerome,Wallace,52,Pharmacist
4000017,Neal,Lawrence,72,Computer support specialist
4000018,Jean,Griffin,45,Childcare worker
4000019,Kristine,Dougherty,63,Financial analyst
4000020,Crystal,Powers,67,Engineering technician
4000021,Alex,May,39,Environmental scientist

```

```

[training@localhost ~]$ hadoop jar /home/training/cust.jar /user/training/custs /user/training/roshan31/cust-part/cust-partitioner
22/09/02 10:18:21 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same
22/09/02 10:18:21 INFO input.FileInputFormat: Total input paths to process : 1
22/09/02 10:18:22 WARN snappy.LoadSnappy: Snappy native library is available
22/09/02 10:18:22 INFO snappy.LoadSnappy: Snappy native library loaded
22/09/02 10:18:22 INFO mapred.JobClient: Running job: job_202209020912_0015
22/09/02 10:18:23 INFO mapred.JobClient: map 0% reduce 0%
22/09/02 10:18:26 INFO mapred.JobClient: map 100% reduce 0%
22/09/02 10:18:30 INFO mapred.JobClient: map 100% reduce 33%
22/09/02 10:18:33 INFO mapred.JobClient: map 100% reduce 66%
22/09/02 10:18:36 INFO mapred.JobClient: map 100% reduce 100%
22/09/02 10:18:36 INFO mapred.JobClient: Job complete: job_202209020912_0015
22/09/02 10:18:36 INFO mapred.JobClient: Counters: 32
22/09/02 10:18:36 INFO mapred.JobClient:   File System Counters
22/09/02 10:18:36 INFO mapred.JobClient:     FILE: Number of bytes read=194431
22/09/02 10:18:36 INFO mapred.JobClient:     FILE: Number of bytes written=1676517
22/09/02 10:18:36 INFO mapred.JobClient:     FILE: Number of read operations=0
22/09/02 10:18:36 INFO mapred.JobClient:     FILE: Number of large read operations=0
22/09/02 10:18:36 INFO mapred.JobClient:     FILE: Number of write operations=0
22/09/02 10:18:36 INFO mapred.JobClient:     HDFS: Number of bytes read=391459
22/09/02 10:18:36 INFO mapred.JobClient:     HDFS: Number of bytes written=872
22/09/02 10:18:36 INFO mapred.JobClient:     HDFS: Number of read operations=2
22/09/02 10:18:36 INFO mapred.JobClient:     HDFS: Number of large read operations=0
22/09/02 10:18:36 INFO mapred.JobClient:     HDFS: Number of write operations=6
22/09/02 10:18:36 INFO mapred.JobClient:   Job Counters
22/09/02 10:18:36 INFO mapred.JobClient:     Launched map tasks=1
22/09/02 10:18:36 INFO mapred.JobClient:     Launched reduce tasks=6
22/09/02 10:18:36 INFO mapred.JobClient:     Data-local map tasks=1

```

HDFS:/user/training/roshan31/c...

localhost.localdomain:50075/browseDirectory.jsp?dir=%2Fuser%2Ftraining%2Froshan31%2Fcust-part%2Fcu...

Most Visited
Hadoop NameNode
Hadoop JobTracker
Hadoop YARN
Hue
HBase Master

Contents of directory /user/training/roshan31/cust-part/cust-partitioner

Goto : /user/training/roshan31/cusl

[Go to parent directory](#)

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
SUCCESS	file	0 KB	1	64 MB	2022-09-02 10:18	rw-r--r--	training	supergroup
logs	dir				2022-09-02 10:18	rw-r--r--	training	supergroup
part-r-00000	file	0.79 KB	1	64 MB	2022-09-02 10:18	rw-r--r--	training	supergroup
part-r-00001	file	0.01 KB	1	64 MB	2022-09-02 10:18	rw-r--r--	training	supergroup
part-r-00002	file	0.01 KB	1	64 MB	2022-09-02 10:18	rw-r--r--	training	supergroup
part-r-00003	file	0.01 KB	1	64 MB	2022-09-02 10:18	rw-r--r--	training	supergroup
part-r-00004	file	0.01 KB	1	64 MB	2022-09-02 10:18	rw-r--r--	training	supergroup
part-r-00005	file	0.01 KB	1	64 MB	2022-09-02 10:18	rw-r--r--	training	supergroup

[Go back to DFS home](#)

HDFS:/user/training/roshan31/c... +

localhost.localdomain:50075/browseBlock.jsp?blockId=-6694312643400980232&blockSize=10&gens

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

File: /user/training/roshan31/cust-part/cust-partitioner/part-r-00001

Goto : /user/training/roshan31/cust go

[Go back to dir listing](#)
[Advanced view/download options](#)

Pilot 209

HDFS:/user/training/roshan31/c... +

localhost.localdomain:50075/browseBlock.jsp?blockId=2089361072181278495&blockSize=12&g

Most Visited ▾ Hadoop NameNode Hadoop JobTracker Hadoop YARN Hue HBase Mast

File: /user/training/roshan31/cust-part/cust-partitioner/part-r-00002

Goto : /user/training/roshan31/cust go

[Go back to dir listing](#)
[Advanced view/download options](#)

Teacher 189

localhost.localdomain:50075/browseBlock.jsp?blockId=-2289328448854701315&

Most Visited ▾ Hadoop NameNode Hadoop JobTracker Hadoop YARN Hue

File: [/user/training/roshan31/cust-part/cust-partitioner/part-r-00003](#)

Goto :

[Go back to dir listing](#)
[Advanced view/download options](#)

Lawyer	201
--------	-----

localhost.localdomain:50075/browseBlock.jsp?blockId=-7022187064034345069&blockSize=15&ge


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





File: [/user/training/roshan31/cust-part/cust-partitioner/part-r-00004](#)

Goto :

[Go back to dir listing](#)
[Advanced view/download options](#)

Pharmacist	213
------------	-----



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

File: [/user/training/roshan31/cust-part/cust-partitioner/part-r-00005](#)

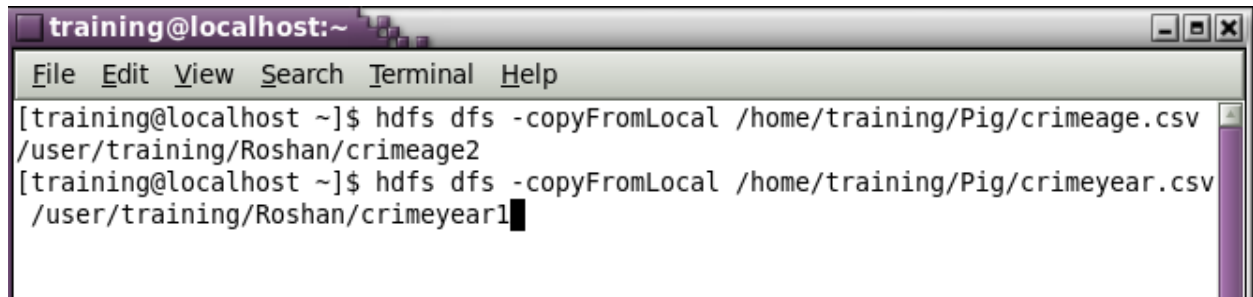
Goto :

[Go back to dir listing](#)
[Advanced view/download options](#)

Carpenter	180
-----------	-----

Assignment : 07

Government Crime Assignment - Pig

A screenshot of a terminal window titled 'training@localhost:~'. The window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The terminal shows two HDFS commands being executed: 'hdfs dfs -copyFromLocal /home/training/Pig/crimeage.csv /user/training/Roshan/crimeage2' and 'hdfs dfs -copyFromLocal /home/training/Pig/crimeyear.csv /user/training/Roshan/crimeyear1'.

```
training@localhost:~  
File Edit View Search Terminal Help  
[training@localhost ~]$ hdfs dfs -copyFromLocal /home/training/Pig/crimeage.csv  
/user/training/Roshan/crimeage2  
[training@localhost ~]$ hdfs dfs -copyFromLocal /home/training/Pig/crimeyear.csv  
/user/training/Roshan/crimeyear1
```

```
[training@localhost ~]$ pig
```

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

```
crimeyear7 = load '/user/training/Roshan/crimeyear2' using PigStorage(',') AS (STATE:chararray,  
CRIMEHEAD:chararray, one:int, two:int, three:int, four:int, five:int, six:int, seven:int, eight:int,  
nine:int, ten:int, eleven:int, twelve:int);
```

```
G = GROUP crimeyear7 BY STATE;
```

```
C = FOREACH G generate group, SUM(crimeyear7.six);
```

```
Dump C;
```

(GOA, 318)
(ASSAM, 16876)
(BIHAR, 23514)
(DELHI, 12414)
(KERALA, 22812)
(ODISHA, 20816)
(PUNJAB, 7764)
(SIKKIM, 78)
(GUJARAT, 36376)
(HARYANA, 13714)
(MANIPUR, 208)
(MIZORAM, 276)
(TRIPURA, 2544)
(NAGALAND, 128)
(JHARKHAND, 8234)
(KARNATAKA, 22070)
(MEGHALAYA, 316)
(RAJASTHAN, 29092)
(CHANDIGARH, 704)
(PUDUCHERRY, 520)
(TAMIL NADU, 18966)
(DAMAN & DIU, 56)
(LAKSHADWEEP, 2)
(MAHARASHTRA, 72394)
(UTTARAKHAND, 4352)
(WEST BENGAL, 44796)
(CHHATTISGARH, 11516)
(D & N HAVELI, 50)
(A & N ISLANDS, 98)
(UTTAR PRADESH, 69440)
(ANDHRA PRADESH, 61320)
(MADHYA PRADESH, 47506)
(JAMMU & KASHMIR, 7792)
(HIMACHAL PRADESH, 2302)
(ARUNACHAL PRADESH, 282)

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

```
crimeage = LOAD '/user/training/Roshan/crimeage' using PigStorage(',') AS (STATE1:chararray,
Crimehead1:chararray, Male_Below_18_Years:int, Female_Below_18_Years:int,
Male_Between_18_30_Years:int, Female_Between_18_30_Years:int,
Male_Between_30_45_Years:int, Female_Between_30_45_Years:int,
Male_Between_45_60_Years:int, Female_Between_45_60_Years:int,
Male_Above_60_Years:int, Female_Above_60_Years:int, Total_Male:int, Total_Female:int,
Grand_Total:int);
```

```
G = GROUP crimeage ALL;
```

```
C = FOREACH G generate group, SUM(crimeage.Female_Between_18_30_Years);
```

```
Job Stats (time in seconds):
JobId  Maps    Reduces  MaxMapTime    MinMapTime    AvgMapTime    MaxReduceTime    MinReduceTime    AvgReduceTime    Alias    Feature Outputs
job_202209080932_0018  1      1      1      1      1      2      2      2      C,G,crimeage  GROUP_BY,COMBINER  hdfs://0.0.0.0:8020/tmp/temp131
3013100/tmp-1449421197,

Input(s):
Successfully read 421 records (38755 bytes) from: "/user/training/Roshan/crimeage2"

Output(s):
Successfully stored 1 records (20 bytes) in: "hdfs://0.0.0.0:8020/tmp/temp1319013100/tmp-1449421197"

Counters:
Total records written : 1
Total bytes written : 20
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_202209080932_0018

2022-09-08 11:51:59,144 [main] WARN  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_
CONVERSION_FAILED 176 time(s).
2022-09-08 11:51:59,144 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-09-08 11:51:59,148 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-09-08 11:51:59,148 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(all,86071)
```

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

```
crimyear = LOAD '/user/training/Roshan/crimyear' using PigStorage(',') AS (STATE:chararray, CRIMEHEAD:chararray, one:int, two:int, three:int, four:int, five:int, six:int, seven:int, eight:int, nine:int, ten:int, eleven:int, twelve:int);
```

```
rape_case = filter crimyear by CRIMEHEAD matches 'RAPE';
```

```
g = FOREACH rape_case generate $0,$14;
```

```
c = ORDER g by $1 desc;
```

```
r = LIMIT c 1;
```

```
grunt> crimyear = LOAD '/user/training/Roshan/crimyear' using PigStorage(',') AS (STATE:chararray, CRIMEHEAD:chararray, one:int, two:int, three:int, four:int, five:int, six:int, seven:int, eight:int, nine:int, ten:int, eleven:int, twelve:int);  
grunt> rape_case = filter crimyear by CRIMEHEAD matches 'RAPE';
```

```
Output(s):  
Successfully stored 1 records (25 bytes) in: "hdfs://0.0.0.0:8020/tmp/temp235804665/tmp808456980"
```

```
Counters:  
Total records written : 1  
Total bytes written : 25  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0
```

```
Job DAG:  
job_202209092317_0004 -> job_202209092317_0005,  
job_202209092317_0005 -> job_202209092317_0006,  
job_202209092317_0006
```

```
2022-09-09 23:34:47,585 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!  
2022-09-09 23:34:47,588 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1  
2022-09-09 23:34:47,589 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1  
(MADHYA PRADESH,4822)  
grunt>
```

4. Find crime wise which state has the highest count.

```
crimeage = LOAD '/user/training/Roshan/crimeage' using PigStorage(',') AS
(STATE1:chararray, Crimehead1:chararray, Male_Below_18_Years:int,
Female_Below_18_Years:int, Male_Between_18_30_Years:int,
Female_Between_18_30_Years:int, Male_Between_30_45_Years:int,
Female_Between_30_45_Years:int, Male_Between_45_60_Years:int,
Female_Between_45_60_Years:int, Male_Above_60_Years:int,
Female_Above_60_Years:int, Total_Male:int, Total_Female:int, Grand_Total:int);
```

```
count = FOREACH crimeage generate $0, $1, $14;
```

```
count2 = group count all;
```

```
max = foreach count2 generate MAX(count.$2) as cr;
```

```
s_count = filter count by $2 == max.cr;
```

```
grunt> crimeage = LOAD '/user/training/Roshan/crimeage' using PigStorage(',') AS (STATE1:chararray, Crimehead1:chararray, Male_Below_18_Years:int, Female_Below_18_Years:int, Male_Between_18_30_Years:int, Female_Between_18_30_Years:int, Male_Between_30_45_Years:int, Female_Between_30_45_Years:int, Male_Between_45_60_Years:int, Female_Between_45_60_Years:int, Male_Above_60_Years:int, Female_Above_60_Years:int, Total_Male:int, Total_Female:int, Grand_Total:int);
grunt>
grunt> count = FOREACH crimeage generate $0, $1, $14;
grunt> count2 = group count all;
grunt> max = foreach count2 generate MAX(count.$2) as cr;
grunt> s_count = filter count by $2 == max.cr;
grunt> dump s_count;
2022-09-09 23:47:55,837 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: GROUP BY,FILTER
2022-09-09 23:47:55,848 [main] INFO org.apache.pig.newplan.logical.rules.ColumnPruneVisitor - Columns pruned for crimeage: $2, $3, $4, $5, $6, $7, $8, $9, $10, $11, $12, $13
2022-09-09 23:47:55,862 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-09-09 23:47:55,865 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - Choosing to move algebraic foreach to combiner
2022-09-09 23:47:55,866 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 3
2022-09-09 23:47:55,866 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - Merged 1 map-reduce splitters.
2022-09-09 23:47:55,866 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - Merged 1 out of total 3 MR operators.
2022-09-09 23:47:55,866 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 2
2022-09-09 23:47:55,867 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the job
2022-09-09 23:47:55,867 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0
2022-09-09 23:47:55,867 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - creating jar file Job3394041714363241732.jar
2022-09-09 23:47:57,698 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - jar file Job3394041714363241732.jar created
2022-09-09 23:47:57,714 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) waiting for submission.
2022-09-09 23:47:57,717 [Thread-94] WARN org.apache.hadoop.mapred.JobClient - Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.
2022-09-09 23:47:57,946 [Thread-94] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-09-09 23:47:57,946 [Thread-94] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2022-09-09 23:47:57,946 [Thread-94] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths (combined) to process : 1
```

```
Output(s):
Successfully stored 1 records (55 bytes) in: "hdfs://0.0.0.0:8020/tmp/temp235804665/tmp831506891"
```

```
Counters:
Total records written : 1
Total bytes written : 55
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0
```

```
Job DAG:
job_202209092317_0008 -> job_202209092317_0009,
job_202209092317_0009
```

```
2022-09-09 23:48:15,682 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-09-09 23:48:15,684 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2022-09-09 23:48:15,685 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(UTTAR PRADESH,TOTAL CRIMES AGAINST WOMEN,77745)
grunt>
```