

**CHAROTAR UNIVERSITY OF SCIENCE & TECHNOLOGY**  
**FACULTY OF TECHNOLOGY AND ENGINEERING**

DEPSTAR

CS442 - Data Science Analytics

Practical Exam

**18DCS007**

**AIM:**

To develop a MapReduce application and implement a program that analyzes Earth Quake data.

**CODE:**

```
package hadoop;
```

```
import java.util.*;
```

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

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

```
import org.apache.hadoop.fs.Path;
```

```
import org.apache.hadoop.conf.*;
```

```
import org.apache.hadoop.io.*;
```

```
import org.apache.hadoop.mapred.*;
```

```
import org.apache.hadoop.util.*;
```

```
public class ProcessUnits {
```

```
    //Mapper class
```

```
    public static class E_EMapper extends MapReduceBase implements
```

```
        Mapper<LongWritable ,/*Input key Type */
```

```
        Text,          /*Input value Type*/
```

```
        Text,          /*Output key Type*/
```

```
IntWritable>    /*Output value Type*/
```

```
{
```

```
    //Map function
```

```
    public void map(LongWritable key, Text value,
```

```
    OutputCollector<Text, IntWritable> output,
```

```
    Reporter reporter) throws IOException {
```

```
        String line = value.toString();
```

```
        String lasttoken = null;
```

```
        StringTokenizer s = new StringTokenizer(line, "\\t");
```

```
        String year = s.nextToken();
```

```
        while(s.hasMoreTokens()) {
```

```
            lasttoken = s.nextToken();
```

```
        }
```

```
        int avgprice = Integer.parseInt(lasttoken);
```

```
        output.collect(new Text(year), new IntWritable(avgprice));
```

```
    }
```

```
}
```

```
//Reducer class
```

```
public static class E_EReduce extends MapReduceBase implements Reducer< Text, IntWritable, Text,
IntWritable > {
```

```
    //Reduce function
```

```
    public void reduce( Text key, Iterator <IntWritable> values,
```

```
    OutputCollector<Text, IntWritable> output, Reporter reporter) throws IOException {
```

```
        int maxavg = 30;
```

```
        int val = Integer.MIN_VALUE;
```

```
        while (values.hasNext()) {  
            if((val = values.next().get())>maxavg) {  
                output.collect(key, new IntWritable(val));  
            }  
        }  
    }  
}
```

//Main function

```
public static void main(String args[])throws Exception {  
    JobConf conf = new JobConf(ProcessUnits.class);  
  
    conf.setJobName("max_earthquakeunits");  
    conf.setOutputKeyClass(Text.class);  
    conf.setOutputValueClass(IntWritable.class);  
    conf.setMapperClass(E_EMapper.class);  
    conf.setCombinerClass(E_EReduce.class);  
    conf.setReducerClass(E_EReduce.class);  
    conf.setInputFormat(TextInputFormat.class);  
    conf.setOutputFormat(TextOutputFormat.class);  
  
    FileInputFormat.setInputPaths(conf, new Path(args[0]));  
    FileOutputFormat.setOutputPath(conf, new Path(args[1]));  
  
    JobClient.runJob(conf);  
}
```

## OUTPUT:

### DATASET

1965	6
1965	6
1965	6
1965	6
1965	6
1965	7
1965	6
1965	6
1965	6
1965	6
1965	9
1965	6
1965	6
1965	6
1965	6
1965	6
1965	6
1965	6

### HADOOP STEP IMAGES

```
hadoopuser@virtual-machine-3f773ebc:~/apache-hive-3.1.2-bin/bin$ hdfs dfs -mkdir /BDA/EarthquakeMap
2021-11-19 04:21:55,720 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoopuser@virtual-machine-3f773ebc:~/apache-hive-3.1.2-bin/bin$
```

```
hadoopuser@virtual-machine-3f773ebc:~/exam$
hadoopuser@virtual-machine-3f773ebc:~/exam$ SHADOOP_HOME/bin/hadoop fs -put /home/hadoopuser/hadoop/earthquake.txt /BDA/EarthquakeMap
2021-11-19 04:23:39,153 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoopuser@virtual-machine-3f773ebc:~/exam$ SHADOOP_HOME/bin/hadoop fs -ls /BDA/EarthquakeMap
2021-11-19 04:23:50,314 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 1 hadoopuser supergroup 376 2021-11-19 04:23 /BDA/EarthquakeMap/earthquake.txt
```

```
hadoopuser@virtual-machine-3f773ebc:~/exam$ SHADOOP_HOME/bin/hadoop jar units.jar hadoop.ProcessUnits /BDA/EarthquakeMap/earthquake.txt /BDA/EarthquakeMap/resul
t
2021-11-19 04:24:14,599 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2021-11-19 04:24:16,716 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2021-11-19 04:24:17,958 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /127.0.0.1:8032
2021-11-19 04:24:18,637 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your appli
cation with ToolRunner to remedy this.
2021-11-19 04:24:18,742 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoopuser/.staging/job_1636692772715_00
08
2021-11-19 04:24:20,336 INFO mapred.FileInputFormat: Total input files to process : 1
2021-11-19 04:24:21,481 INFO mapreduce.JobSubmitter: number of splits:2
2021-11-19 04:24:22,626 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1636692772715_0008
2021-11-19 04:24:22,626 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-11-19 04:24:23,340 INFO conf.Configuration: resource-types.xml not found
2021-11-19 04:24:23,340 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-11-19 04:24:23,621 INFO impl.YarnClientImpl: Submitted application application_1636692772715_0008
2021-11-19 04:24:23,929 INFO mapreduce.Job: The url to track the job: http://virtual-machine-3f773ebc.localdomain:8088/proxy/application_1636692772715_0008/
2021-11-19 04:24:23,938 INFO mapreduce.Job: Running job: job_1636692772715_0008
2021-11-19 04:24:46,411 INFO mapreduce.Job: Job job_1636692772715_0008 running in uber mode : false
2021-11-19 04:24:46,415 INFO mapreduce.Job: map 0% reduce 0%
2021-11-19 04:25:12,835 INFO mapreduce.Job: map 100% reduce 0%
2021-11-19 04:25:23,141 INFO mapreduce.Job: map 100% reduce 100%
2021-11-19 04:25:25,282 INFO mapreduce.Job: Job job_1636692772715_0008 completed successfully
2021-11-19 04:25:25,453 INFO mapreduce.Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=6
    FILE: Number of bytes written=818150
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=776
    HDFS: Number of bytes written=0
    HDFS: Number of read operations=11
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Launched map tasks=2
    Launched reduce tasks=1
    Data-local map tasks=2
    Total time spent by all maps in occupied slots (ms)=45700
    Total time spent by all reduces in occupied slots (ms)=8297
    Total time spent by all map tasks (ms)=45700
    Total time spent by all reduce tasks (ms)=8297
    Total vcore-milliseconds taken by all map tasks=45700
    Total vcore-milliseconds taken by all reduce tasks=8297
    Total megabyte-milliseconds taken by all map tasks=46796800
    Total megabyte-milliseconds taken by all reduce tasks=8496128
  Map-Reduce Framework
    Total megabyte-milliseconds taken by all reduce tasks=8496128
```

```
Map-Reduce Framework
  Map input records=47
  Map output records=47
  Map output bytes=423
  Map output materialized bytes=12
  Input split bytes=212
  Combine input records=47
  Combine output records=0
  Reduce input groups=0
  Reduce shuffle bytes=12
  Reduce input records=0
  Reduce output records=0
  Spilled Records=0
  Shuffled Maps =2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=741
  CPU time spent (ms)=6750
  Physical memory (bytes) snapshot=816361472
  Virtual memory (bytes) snapshot=7618895872
  Total committed heap usage (bytes)=710934528
  Peak Map Physical memory (bytes)=309927936
  Peak Map Virtual memory (bytes)=2537377792
  Peak Reduce Physical memory (bytes)=201564160
  Peak Reduce Virtual memory (bytes)=2544300032
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=564
  File Output Format Counters
    Bytes Written=0
hadoopuser@virtual-machine-3f773ebc:~/exam$
```

```
Bytes Written=0
hadoopuser@virtual-machine-3f773ebc:~/exam$ SHADOOP_HOME/bin/hadoop fs -ls /BDA/EarthquakeMap/result
2021-11-19 04:28:11,895 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 hadoopuser supergroup 0 2021-11-19 04:25 /BDA/EarthquakeMap/result/_SUCCESS
-rw-r--r-- 1 hadoopuser supergroup 0 2021-11-19 04:25 /BDA/EarthquakeMap/result/part-00000
hadoopuser@virtual-machine-3f773ebc:~/exam$ SHADOOP_HOME/bin/hadoop fs -cat /BDA/EarthquakeMap/result/part-00000
```

# Browse Directory

/BDA/EarthquakeMap

Go!

Show

>5

entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
<input type="checkbox"/>	-rw-r--r--	<a href="#">hadoopuser</a>	<a href="#">supergroup</a>	376 B	Nov 19 04:23	<a href="#">1</a>	128 MB	<a href="#">earthquake.txt</a>
<input type="checkbox"/>	drwxr-xr-x	<a href="#">hadoopuser</a>	<a href="#">supergroup</a>	0 B	Nov 19 04:25	<a href="#">0</a>	0 B	<a href="#">result</a>

Showing 1 to 2 of 2 entries

Previous

1

Next

Hadoop, 2021.

# Browse Directory

/BDA/EarthquakeMap/result

Go!

Show

>5

entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
<input type="checkbox"/>	-rw-r--r--	<a href="#">hadoopuser</a>	<a href="#">supergroup</a>	0 B	Nov 19 04:25	<a href="#">1</a>	128 MB	<a href="#">_SUCCESS</a>
<input type="checkbox"/>	-rw-r--r--	<a href="#">hadoopuser</a>	<a href="#">supergroup</a>	0 B	Nov 19 04:25	<a href="#">1</a>	128 MB	<a href="#">part-00000</a>

Showing 1 to 2 of 2 entries

Previous

1

Next

Hadoop, 2021.