Name - Prathamesh Chaudhari
Roll.no - TI10

PartA: Assignment No2

Aim: Design a distributed application using MapReduce which processes log file of asystem.List out users who have logged for maximum period on the system.

Name of input file is access_log_short.csv

PARTA

1. Open Eclipse> File > New > Java Project >(Name it – MRProgramsDemo) > Next>Click on Libraries Tab>Click on Add External JARS tab

   jar FILE LOCATION
   /usr/lib/Hadoop ->select all jar files
   /usr/lib/Hadoop/client ->select all jar files

2. Right Click > New > Package (Name it - mrLogFile_demo > Finish.
3. Right Click on mrLogFile_demo Package > New > Class (Name it – UserLogDriver).

# Add following code in that class

```
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapred.*;

public class UserLogDriver {
      public static void main(String[] args) {
            JobClient my_client = new JobClient();
            // Create a configuration object for the job
            JobConf job_conf = new JobConf(UserLogDriver.class);

            // Set a name of the Job
            job_conf.setJobName("MaxLoggedUsers");

            // Specify data type of output key and value
            job_conf.setOutputKeyClass(Text.class);
            job_conf.setOutputValueClass(IntWritable.class);

            // Specify names of Mapper and Reducer Class
            job_conf.setMapperClass(UserLogMapper.class);


            job_conf.setReducerClass(UserLogReducer .class);

            // Specify formats of the data type of Input and output
            job_conf.setInputFormat(TextInputFormat.class);
            job_conf.setOutputFormat(TextOutputFormat.class);

            // Set input and output directories using command line arguments,
            //arg[0] = name of input directory on HDFS, and arg[1] =  name of
output directory to be created to store the output file.

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

            my_client.setConf(job_conf);
            try {
                  // Run the job
                  JobClient.runJob(job_conf);
            } catch (Exception e) {
                  e.printStackTrace();
            }
      }
}
```

# Save the file

4. Right Click on mrLogFile_demo Package > New > Class (Name it – UserLogReducer).

```
import java.io.IOException;
import java.util.*;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.*;

public class UserLogReducer  extends MapReduceBase implements Reducer<Text,
IntWritable, Text, IntWritable> {

      public void reduce(Text t_key, Iterator<IntWritable> values,
OutputCollector<Text,IntWritable> output, Reporter reporter) throws IOException
{
            Text key = t_key;
            int frequencyForUser = 0;
            while (values.hasNext()) {
                  // replace type of value with the actual type of our value
                  IntWritable value = (IntWritable) values.next();
                  frequencyForUser += value.get();

            }
            output.collect(key, new IntWritable(frequencyForUser));
      }
}
```

# Save the file

5. **Right Click on mrLogFile_demo Package > New > Class (Name it – `UserLogMapper`).**

# Add following code in that class

```
package MRLogFile;
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapred.*;

public class UserLogMapper extends MapReduceBase implements Mapper<LongWritable,
Text, Text, IntWritable> {
	private final static IntWritable one = new IntWritable(1);

	public void map(LongWritable key, Text value, OutputCollector<Text,
IntWritable> output, Reporter reporter) throws IOException {

		String valueString = value.toString();
		String[] SingleUserData = valueString.split("-");
		output.collect(new Text(SingleUserData[0]), one);
	}
}
```

# Save the file

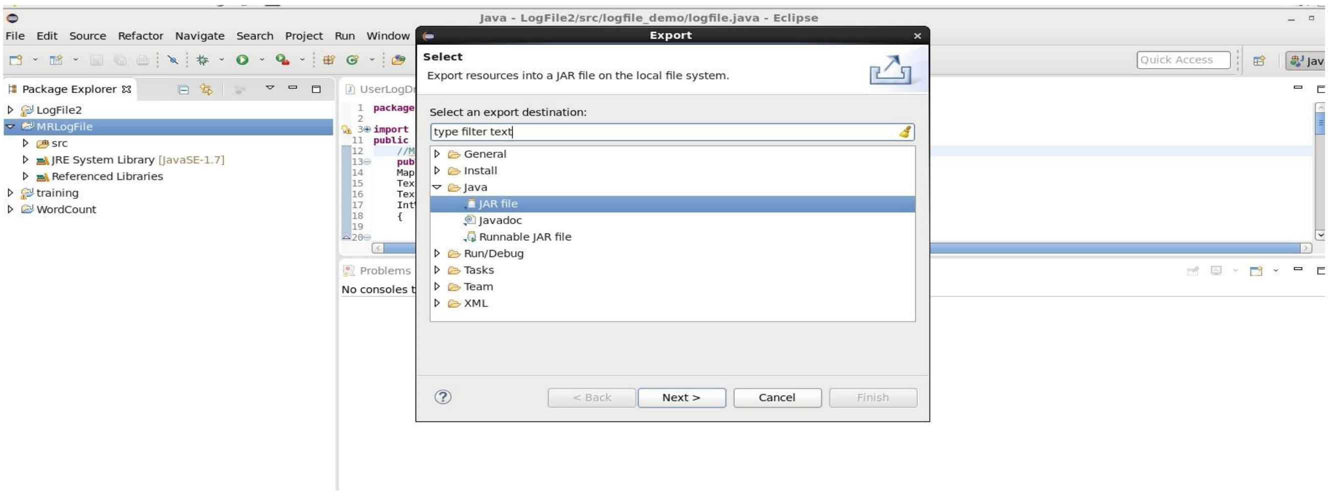
PART B

Create .jar file for your program execution :

Make a jar file

In eclipse Right click on MRLogFile Project > then select Export> Click on Java>JAR Files>Click on Next>then select export destination for JAR file as /home/Cloudera/MRlogFile.jar>Finish

*MRLogFile.jar file will get created in your /home/Cloudera/ folder



Export

Select

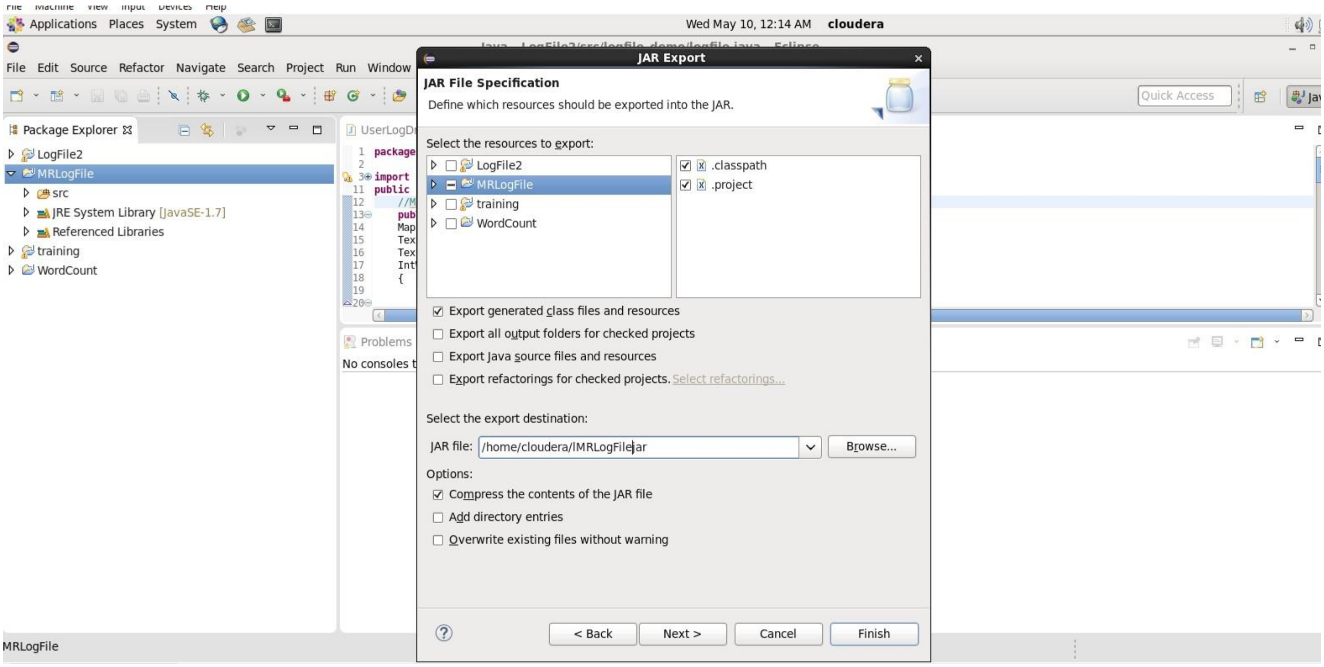
Export resources into a JAR file on the local file system.

Select an export destination:

type filter text

- General
- Install
- Java
  - JAR file
  - Javadoc
  - Runnable JAR file
- Run/Debug
- Tasks
- Team
- XML

Java - LogFile2/src/logfile_demo/logfile.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window
Package Explorer
LogFile2
MRLogFile
src
JRE System Library [JavaSE-1.7]
Referenced Libraries
training
WordCount
Problems
No consoles t
Quick Access
< Back
Next >
Cancel
Finish



JAR Export

JAR File Specification

Define which resources should be exported into the JAR.

Select the resources to export:

- □ LogFile2
- ▣ MRLogFile
- □ training
- □ WordCount

☑ .classpath
☑ .project

☑ Export generated class files and resources
□ Export all output folders for checked projects
□ Export Java source files and resources
□ Export refactorings for checked projects. Select refactorings...

Select the export destination:

JAR file: /home/cloudera/MRLogFile.jar

Options:

☑ Compress the contents of the JAR file
□ Add directory entries
□ Overwrite existing files without warning

PART C:

# Open terminal

#Check for present working Directory

```
[cloudera@quickstart ~]$ pwd
/home/cloudera
```

#Create inputfoder with name MRinputfolder1

```
[cloudera@quickstart ~]$ hdfs dfs -mkdir /MRinputfolder1
```

```
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 21 items
drwxr-xr-x   - cloudera supergroup          0 2023-05-10 00:22 /MRInputfolder
drwxr-xr-x   - cloudera supergroup          0 2023-05-10 00:29 /MRInputfolder1
drwxr-xr-x   - cloudera supergroup          0 2023-05-10 00:38
/MRoutputfolder1
drwxrwxrwx   - hdfs     supergroup          0 2017-10-23 09:15 /benchmarks
drwxr-xr-x   - hbase    supergroup          0 2023-05-10 00:02 /hbase
drwxr-xr-x   - cloudera supergroup          0 2023-05-06 01:27 /inputfolder
drwxr-xr-x   - cloudera supergroup          0 2023-05-07 23:02 /inputfolder1
drwxr-xr-x   - cloudera supergroup          0 2023-05-08 01:45 /inputfolder5
drwxr-xr-x   - cloudera supergroup          0 2023-05-08 03:10 /inputfolder8
drwxr-xr-x   - cloudera supergroup          0 2023-05-08 03:13 /inputfolder9
drwxr-xr-x   - cloudera supergroup          0 2023-05-08 03:31 /out10
drwxr-xr-x   - cloudera supergroup          0 2023-05-08 03:38 /out11
drwxr-xr-x   - cloudera supergroup          0 2023-05-08 03:50 /out14
drwxr-xr-x   - cloudera supergroup          0 2023-05-07 23:55 /out2
drwxr-xr-x   - cloudera supergroup          0 2023-05-08 03:22 /out9
drwxr-xr-x   - cloudera supergroup          0 2023-05-06 01:28 /outputfolder
drwxr-xr-x   - cloudera supergroup          0 2023-05-07 23:04 /outputfolder1
drwxr-xr-x   - solr     solr                0 2017-10-23 09:18 /solr
drwxrwxrwt   - hdfs     supergroup          0 2023-05-05 23:26 /tmp
drwxr-xr-x   - hdfs     supergroup          0 2017-10-23 09:17 /user
drwxr-xr-x   - hdfs     supergroup          0 2017-10-23 09:17 /var
```

```
[cloudera@quickstart ~]$ hdfs dfs -put
/home/cloudera/access_log_short.txt /MRInputfolder1

[cloudera@quickstart ~]$ hdfs dfs -cat
/MRInputfolder1/access_log_short.txt

[cloudera@quickstart ~]$ hadoop jar /home/cloudera/MRLogFile.jar
mrLogFile_demo.UserLogDriver /MRInputfolder1/access_log_short.txt
/MRoutputfolder1
```

```
23/05/10 00:38:06 INFO client.RMProxy: Connecting to ResourceManager at
/0.0.0.0:8032
23/05/10 00:38:06 INFO client.RMProxy: Connecting to ResourceManager at
/0.0.0.0:8032
```

```
23/05/10 00:38:07 WARN mapreduce.JobResourceUploader: Hadoop command-line
option parsing not performed. Implement the Tool interface and execute your
application with ToolRunner to remedy this.
23/05/10 00:38:07 INFO mapred.FileInputFormat: Total input paths to process :
1
23/05/10 00:38:07 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
      at java.lang.Object.wait(Native Method)
      at java.lang.Thread.join(Thread.java:1281)
      at java.lang.Thread.join(Thread.java:1355)
      at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputSt
ream.java:967)
      at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.j
ava:705)
      at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:8
94)
23/05/10 00:38:07 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
      at java.lang.Object.wait(Native Method)
      at java.lang.Thread.join(Thread.java:1281)
      at java.lang.Thread.join(Thread.java:1355)
      at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputSt
ream.java:967)
      at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.j
ava:705)
      at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:8
94)
23/05/10 00:38:07 INFO mapreduce.JobSubmitter: number of splits:2
23/05/10 00:38:08 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1683702103820_0001
23/05/10 00:38:08 INFO impl.YarnClientImpl: Submitted application
application_1683702103820_0001
23/05/10 00:38:08 INFO mapreduce.Job: The url to track the job:
http://quickstart.cloudera:8088/proxy/application_1683702103820_0001/
23/05/10 00:38:08 INFO mapreduce.Job: Running job: job_1683702103820_0001
23/05/10 00:38:19 INFO mapreduce.Job: Job job_1683702103820_0001 running in
uber mode : false
23/05/10 00:38:19 INFO mapreduce.Job:  map 0% reduce 0%
23/05/10 00:38:37 INFO mapreduce.Job:  map 100% reduce 0%
23/05/10 00:38:46 INFO mapreduce.Job:  map 100% reduce 100%
23/05/10 00:38:47 INFO mapreduce.Job: Job job_1683702103820_0001 completed
successfully
23/05/10 00:38:47 INFO mapreduce.Job: Counters: 49
      File System Counters
            FILE: Number of bytes read=26793
            FILE: Number of bytes written=484376
            FILE: Number of read operations=0
            FILE: Number of large read operations=0
            FILE: Number of write operations=0
            HDFS: Number of bytes read=147418
            HDFS: Number of bytes written=3838
            HDFS: Number of read operations=9
            HDFS: Number of large read operations=0
            HDFS: Number of write operations=2
      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)=28992
            Total time spent by all reduces in occupied slots (ms)=7394
            Total time spent by all map tasks (ms)=28992
            Total time spent by all reduce tasks (ms)=7394
            Total vcore-milliseconds taken by all map tasks=28992
            Total vcore-milliseconds taken by all reduce tasks=7394
            Total megabyte-milliseconds taken by all map tasks=29687808
            Total megabyte-milliseconds taken by all reduce tasks=7571456
      Map-Reduce Framework
            Map input records=1295
            Map output records=1295
            Map output bytes=24197
            Map output materialized bytes=26799
            Input split bytes=238
            Combine input records=0
            Combine output records=0
            Reduce input groups=227
            Reduce shuffle bytes=26799
            Reduce input records=1295
            Reduce output records=227
            Spilled Records=2590
            Shuffled Maps =2
            Failed Shuffles=0
            Merged Map outputs=2
            GC time elapsed (ms)=311
            CPU time spent (ms)=2690
            Physical memory (bytes) snapshot=556244992
            Virtual memory (bytes) snapshot=4519596032
            Total committed heap usage (bytes)=391979008
      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=147180
      File Output Format Counters
            Bytes Written=3838
[cloudera@quickstart ~]$ hdfs dfs -ls /MRoutputfolder1
Found 2 items
-rw-r--r--   1 cloudera supergroup          0 2023-05-10 00:38
/MRoutputfolder1/_SUCCESS
-rw-r--r--   1 cloudera supergroup       3838 2023-05-10 00:38
/MRoutputfolder1/part-00000
[cloudera@quickstart ~]$ hdfs dfs -cat /MRoutputfolder1/part-00000
10.1.1.236  7
10.1.181.142     14
10.1.232.31      5
10.10.55.142     14
10.102.101.66    1
10.103.184.104   1
10.103.190.81    53
10.103.63.29     1
10.104.73.51     1
10.105.160.183   1
10.108.91.151    1
10.109.21.76     1
10.11.131.40     1
10.111.71.20     8
10.112.227.184   6
10.114.74.30     1
10.115.118.78    1
```

```
10.117.224.230    1
10.117.76.22      12
10.118.19.97      1
10.118.250.30     7
10.119.117.132    23
10.119.33.245     1
10.119.74.120     1
10.12.113.198     2
10.12.219.30      1
10.120.165.113    1
10.120.207.127    4
10.123.124.47     1
10.123.35.235     1
10.124.148.99     1
10.124.155.234    1
10.126.161.13     7
10.127.162.239    1
10.128.11.75      10
10.13.42.232      1
10.130.195.163    8
10.130.70.80      1
10.131.163.73     1
10.131.209.116    5
10.132.19.125     2
10.133.222.184    12
10.134.110.196    13
10.134.242.87     1
10.136.84.60      5
10.14.2.86 8
10.14.4.151       2
10.140.139.116    1
10.140.141.1      9
10.140.67.116     1
10.141.221.57     5
10.142.203.173    7
10.143.126.177    32
10.144.147.8      1
10.15.208.56      1
10.15.23.44       13
10.150.212.239    14
10.150.227.16     1
10.150.24.40      13
10.152.195.138    8
10.153.23.63      2
10.153.239.5      25
10.155.95.124     9
10.156.152.9      1
10.157.176.158    1
10.164.130.155    1
10.164.49.105     8
10.164.95.122     10
10.165.106.173    14
10.167.1.145      19
10.169.158.88     1
10.170.178.53     1
10.171.104.4      1
10.172.169.53     18
10.174.246.84     3
10.175.149.65     1
10.175.204.125    15
10.177.216.164    6
10.179.107.170    2
10.181.38.207     13
10.181.87.221     1
```

```
10.185.152.140    1
10.186.56.126     16
10.186.56.183     1
10.187.129.140    6
10.187.177.220    1
10.187.212.83     1
10.187.28.68      1
10.19.226.186     2
10.190.174.142    10
10.190.41.42      5
10.191.172.11     1
10.193.116.91     1
10.194.174.4      7
10.198.138.192    1
10.199.103.248    2
10.199.189.15     1
10.2.202.135      1
10.200.184.212    1
10.200.237.222    1
10.200.9.128      2
10.203.194.139    10
10.205.72.238     2
10.206.108.96     2
10.206.175.236    1
10.206.73.206     7
10.207.190.45     17
10.208.38.46      1
10.208.49.216     4
10.209.18.39      9
10.209.54.187     3
10.211.47.159     10
10.212.122.173    1
10.213.181.38     7
10.214.35.48      1
10.215.222.114    1
10.216.113.172    48
10.216.134.214    1
10.216.227.195    16
10.217.151.145    10
10.217.32.16      1
10.218.16.176     8
10.22.108.103     4
10.220.112.1      34
10.221.40.89      5
10.221.62.23      13
10.222.246.34     1
10.223.157.186    11
10.225.137.152    1
10.225.234.46     1
10.226.130.133    1
10.229.60.23      1
10.230.191.135    6
10.231.55.231     1
10.234.15.156     1
10.236.231.63     1
10.238.230.235    1
10.239.100.52     1
10.239.52.68      4
10.24.150.4       5
10.24.67.131      13
10.240.144.183    15
10.240.170.50     1
10.241.107.75     1
10.241.9.187      1
```

```
10.243.51.109     5
10.244.166.195    5
10.245.208.15     20
10.246.151.162    3
10.247.111.104    9
10.247.175.65     1
10.247.229.13     1
10.248.24.219     1
10.248.36.117     3
10.249.130.132    3
10.25.132.238     2
10.25.44.247      6
10.250.166.232    1
10.27.134.23      1
10.30.164.32      1
10.30.47.170      8
10.31.225.14      7
10.32.138.48      11
10.32.247.175     4
10.32.55.216      12
10.33.181.9       8
10.34.233.107     1
10.36.200.176     1
10.39.45.70       2
10.39.94.109      4
10.4.59.153       1
10.4.79.47  15
10.41.170.233     9
10.41.40.17       1
10.42.208.60      1
10.43.81.13       1
10.46.190.95      10
10.48.81.158      5
10.5.132.217      1
10.5.148.29       1
10.50.226.223     9
10.50.41.216      3
10.52.161.126     1
10.53.58.58       1
10.54.242.54      10
10.54.49.229      1
10.56.48.40       16
10.59.42.194      11
10.6.238.124      6
10.61.147.24      1
10.61.161.218     1
10.61.23.77       8
10.61.232.147     3
10.62.78.165      2
10.63.233.249     7
10.64.224.191     13
10.66.208.82      2
10.69.20.85       26
10.70.105.238     1
10.70.238.46      6
10.72.137.86      6
10.72.208.27      1
10.73.134.9       4
10.73.238.200     1
10.73.60.200      1
10.73.64.91       1
10.74.218.123     1
10.75.116.199     1
10.76.143.30      1
```

```
10.76.68.178      16
10.78.95.24       8
10.80.10.131      10
10.80.215.116     17
10.81.134.180     1
10.82.30.199      63
10.82.64.235      1
10.84.236.242     1
10.87.209.46      1
10.87.88.214      1
10.88.204.177     1
10.89.178.62      1
10.89.244.42      1
10.94.196.42      1
10.95.136.211     4
10.95.232.88      1
10.98.156.141     1
10.99.228.224     1
[cloudera@quickstart ~]$
```