

Name: Pooja Pathak

Roll No: 14

MSc Part-I Sem-2

Subject: Big Data Technology


Practical 4: Max Temp


File Edit Source Refactor Navigate Search Project Run Window Help		
New	Shift+Alt+N >	Java Project
Open File...		Project...
Close	Ctrl+W	Package
Close All	Shift+Ctrl+W	Class
Save	Ctrl+S	Interface
Save As...		Enum
Save All	Shift+Ctrl+S	Annotation
Revert		Source Folder
Move...		Java Working Set
Rename...	F2	Folder
Refresh	F5	File
Convert Line Delimiters To	>	Untitled Text File
		JUnit Test Case

Name: Pooja Pathak

Roll No: 14

MSc Part-I Sem-2

 Change account settings and status ×

Create a Java Project 


Create a Java project in the workspace or in an external location.


Project name:

☒ Use default location

Location:

JRE

☒ Use an execution environment JRE: 

☐ Use a project specific JRE: 

☐ Use default JRE (currently 'jdk1.7.0_67-cloudera') [Configure JREs...](#)


Project layout


☐ Use project folder as root for sources and class files


☒ Create separate folders for sources and class files [Configure default...](#)

Working sets

☐ Add project to working sets

Working sets: 

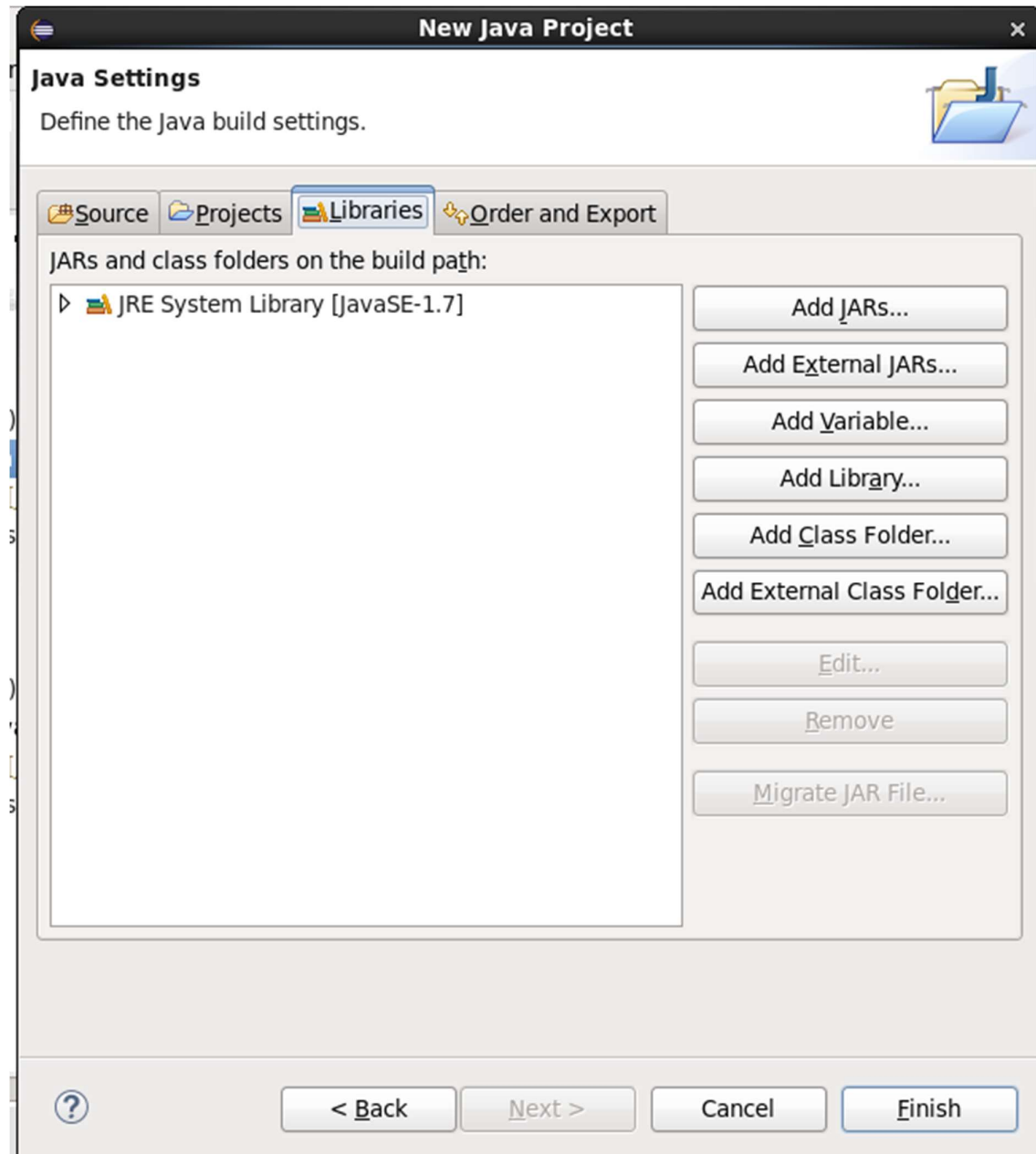
 The default compiler compliance level for the current workspace is 1.6. The new project will use a project specific compiler compliance level of 1.7.



Name: Pooja Pathak

Roll No: 14

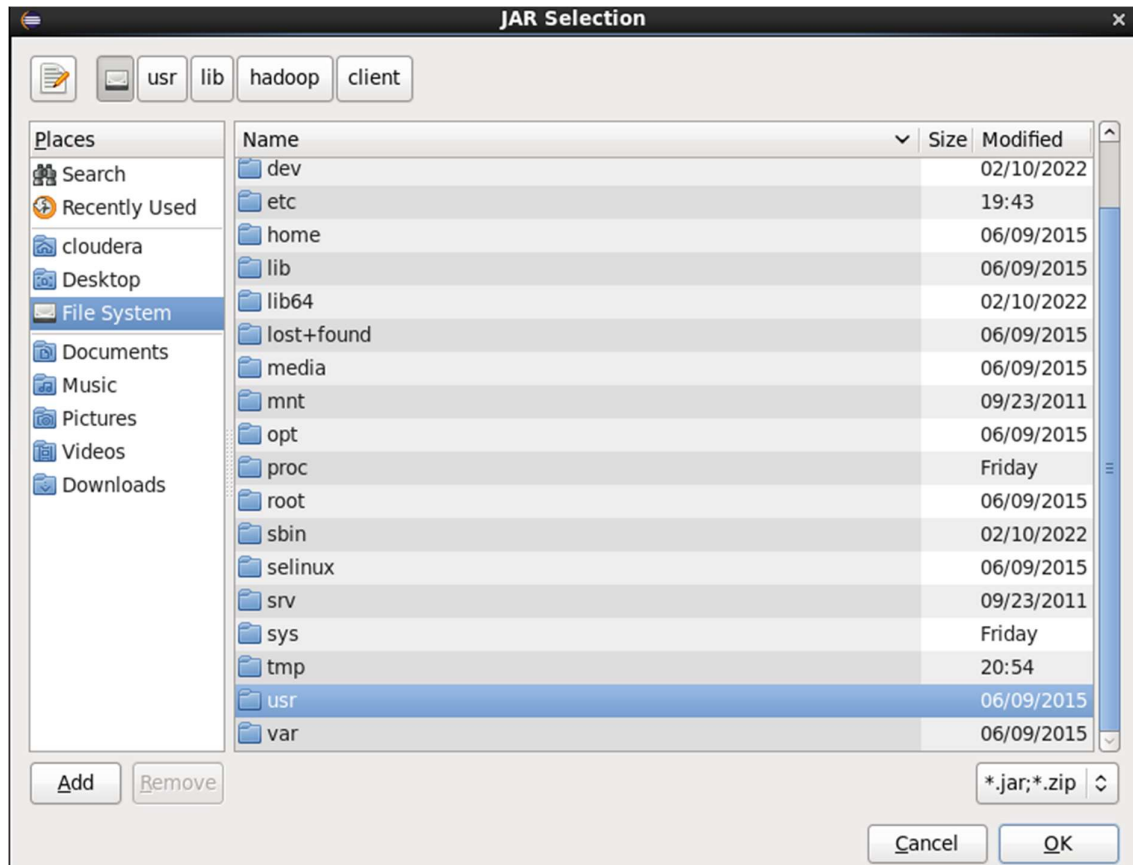
MSc Part-I Sem-2



Name: Pooja Pathak

Roll No: 14

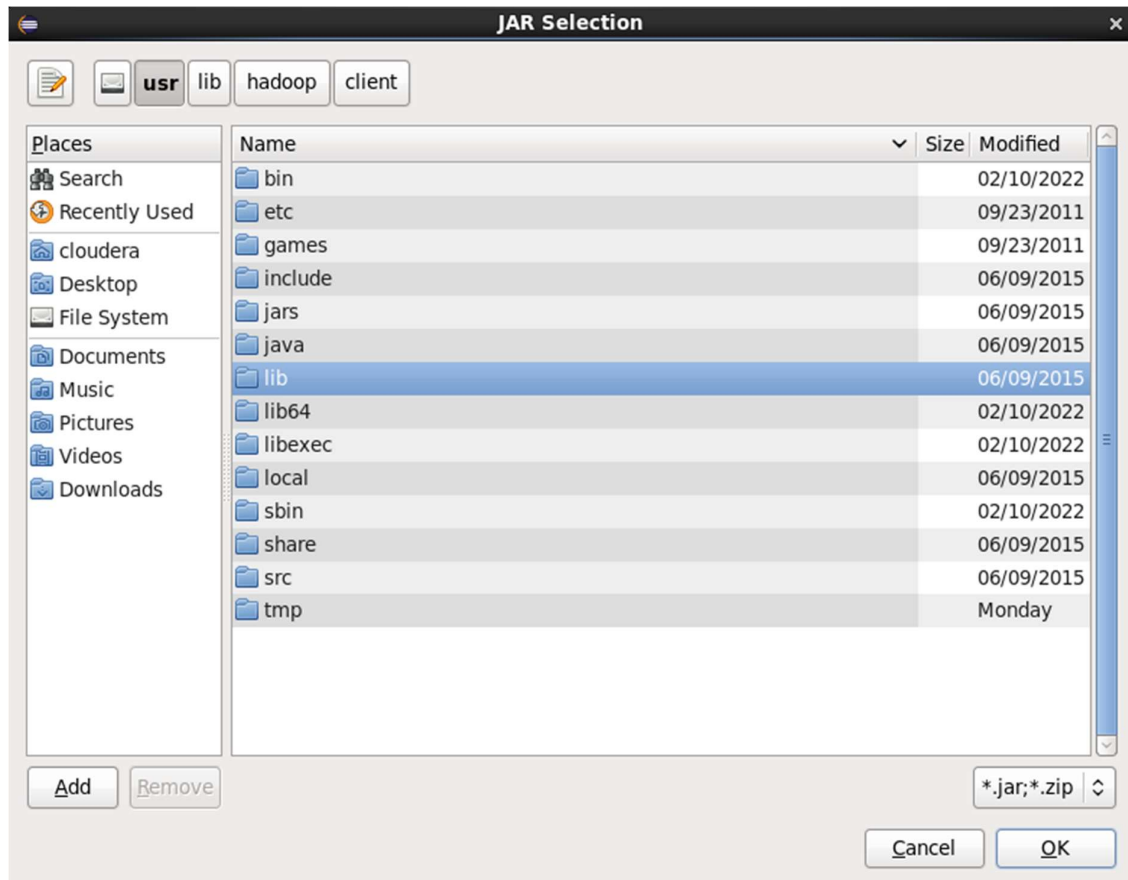
MSc Part-I Sem-2



Name: Pooja Pathak

Roll No: 14

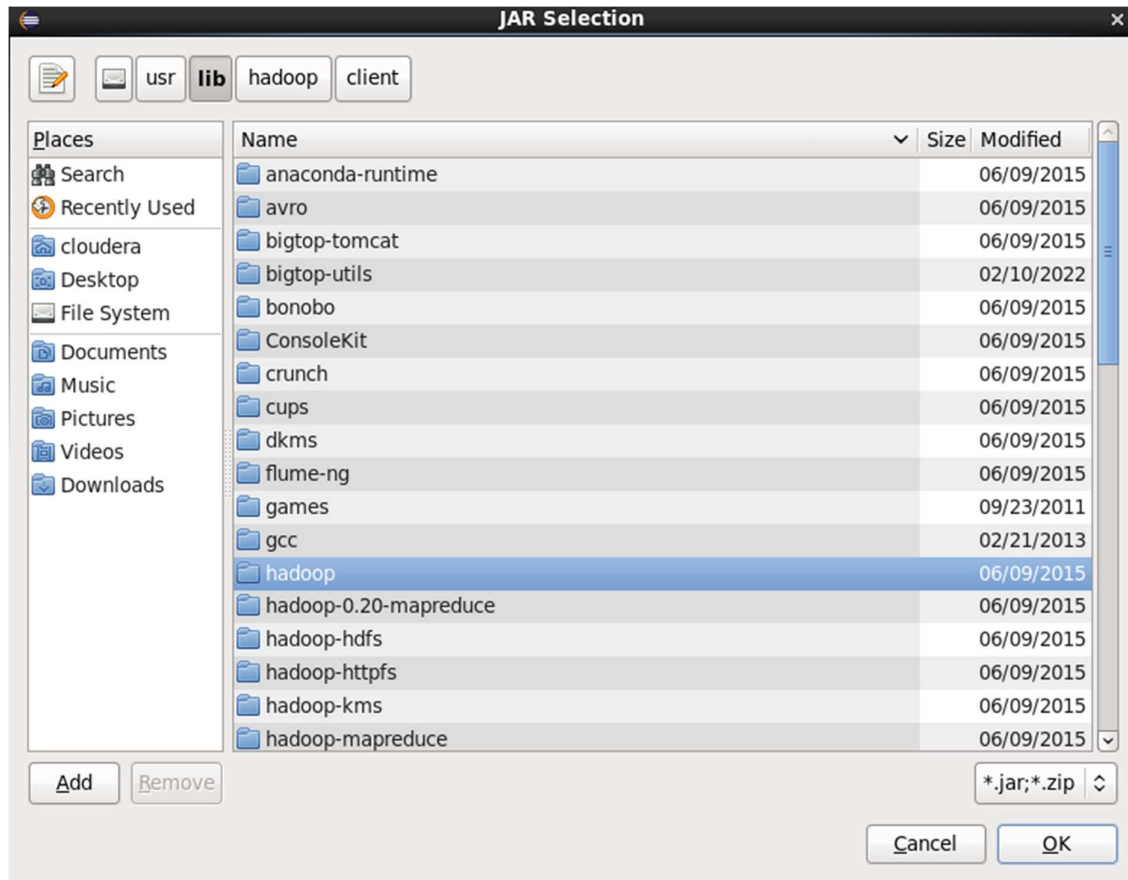
MSc Part-I Sem-2



Name: Pooja Pathak

Roll No: 14

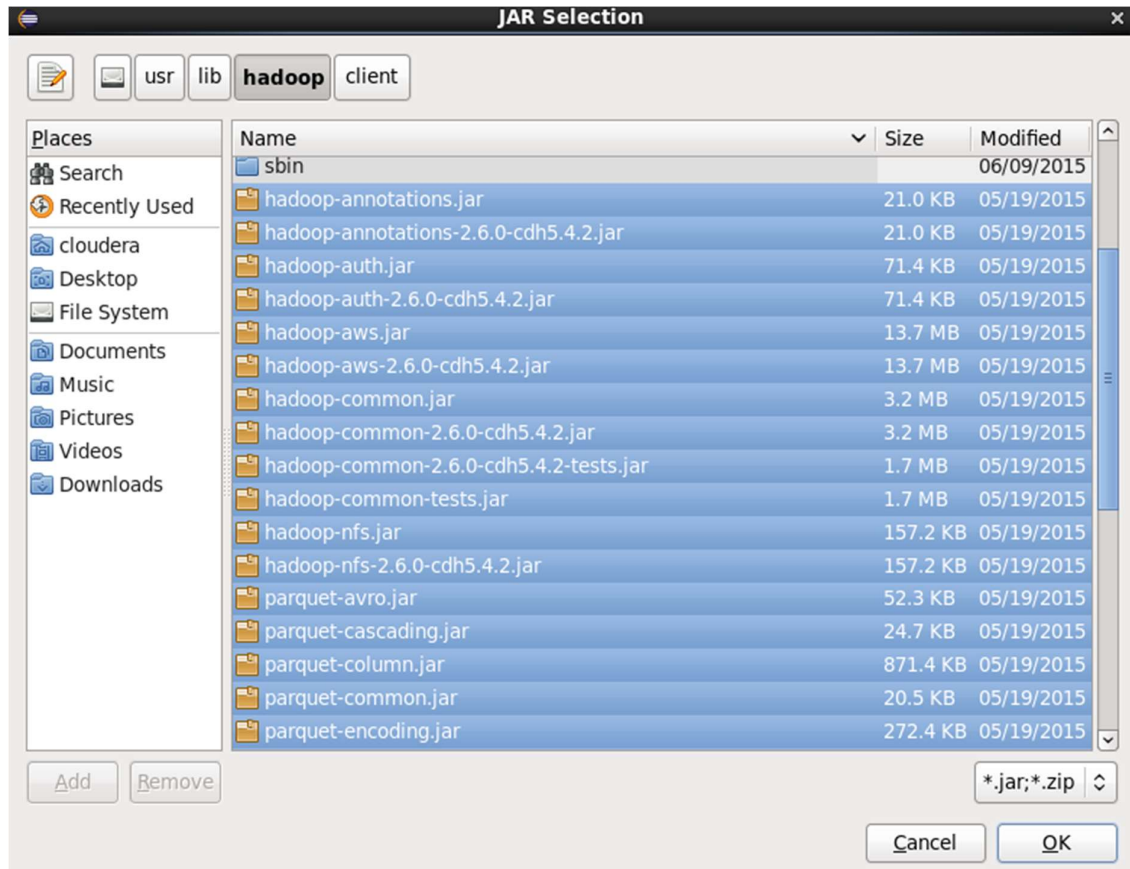
MSc Part-I Sem-2



Name: Pooja Pathak

Roll No: 14

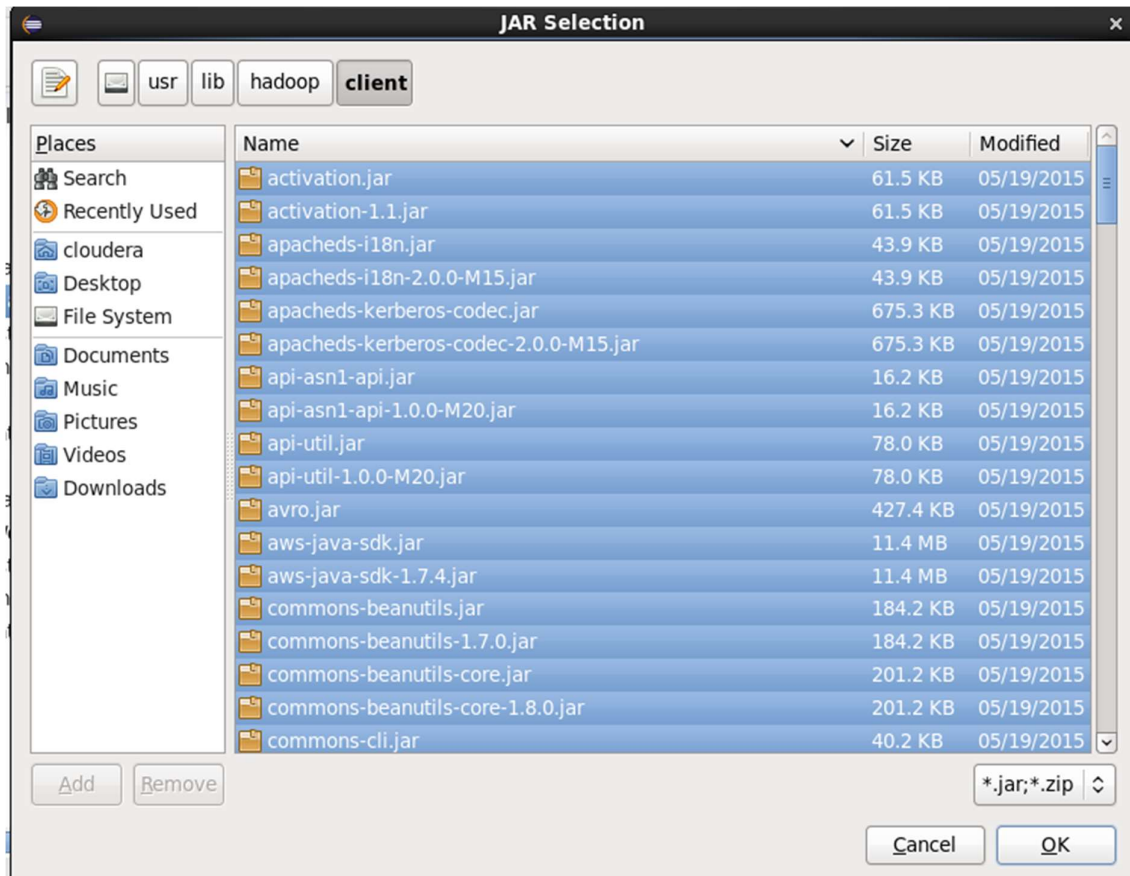
MSc Part-I Sem-2



Name: Pooja Pathak

Roll No: 14

MSc Part-I Sem-2



Name: Pooja Pathak

Roll No: 14

MSc Part-I Sem-2

```
StubDriver.java WordCount.java MaxTemp.java
1 import java.io.IOException;
18 //import MaximumTemp.MaxTemperatureMapper;
19 //import MaxTemp.MaxTemperatureReducer;
20
21 public class MaxTemp {
22     // Mapper
23     /*MaxTemperatureMapper class is static
24     * and extends Mapper abstract class
25     * having four Hadoop generics type
26     * LongWritable, Text, Text, Text.
27     */
28     public static class MaxTemperatureMapper extends
29     Mapper<LongWritable, Text, Text, IntWritable> {
30         // the data in our data set with
31         // this value is inconsistent data
32         //public static final int MISSING = 9999;
33
34         public void map(LongWritable key, Text value, Context context)
35         throws IOException, InterruptedException {
36             String line=value.toString();
37             String year=line.substring(15,19);
38             int airtemp;
39             if(line.charAt(87)=='+')
40             {
41                 airtemp=Integer.parseInt(line.substring(88,92));
42             }
43             else
44                 airtemp=Integer.parseInt(line.substring(87,92));
45             String q=line.substring(92,93);
46             if(airtemp!=9999 && q.matches("[01459]"))
47             {
48
49                 context.write(new Text(year), new IntWritable(airtemp));
50             }
51         }
52     }
53     // Reducer
54     /*MaxTemperatureReducer class is static
55     and extends Reducer abstract class
56     having four Hadoop generics type
57     Text, Text, Text, Text.
58     */
59     public static class MaxTemperatureReducer extends
60     Reducer<Text, IntWritable, Text, IntWritable> {
61         /**
62         * @method reduce
```

Name: Pooja Pathak

Roll No: 14

MSc Part-I Sem-2

```
62  * @method reduce
63  * This method takes the input as key and
64  * list of values pair from the mapper,
65  * it does aggregation based on keys and
66  * produces the final context.
67  */
68  public void reduce(Text key, Iterable<IntWritable> values, Context context)
69  throws IOException, InterruptedException {
70  int maxvalue= Integer.MIN_VALUE;
71  for (IntWritable value : values) {
72  maxvalue=Math.max(maxvalue, value.get());
73  }
74  context.write(key, new IntWritable(maxvalue));
75  }
76  }
77
78  /**
79  * @method main
80  * This method is used for setting
81  * all the configuration properties.
82  * It acts as a driver for map-reduce
83  * code.
84  */
85  public static void main(String[] args) throws Exception {
86  // reads the default configuration of the
87  // cluster from the configuration XML files
88  Configuration conf = new Configuration();
89  // Initializing the job with the
90  // default configuration of the cluster
91  // Job job = new Job(conf, "weather example");
92  Job job = Job.getInstance(conf, "weather example");
93  // Assigning the driver class name
94  job.setJarByClass(MaxTemp.class);
95  // Key type coming out of mapper
96  // job.setMapOutputKeyClass(Text.class);
97  // value type coming out of mapper
98  // job.setMapOutputValueClass(Text.class);
99  // Defining the mapper class name
100  job.setMapperClass(MaxTemperatureMapper.class);
101  // Defining the reducer class name
102  job.setReducerClass(MaxTemperatureReducer.class);
103  // Defining input Format class which is
104  // responsible to parse the dataset
105  // into a key value pair
106  job.setInputFormatClass(TextInputFormat.class);
107  // Defining output Format class which is
108
109  job.setInputFormatClass(TextInputFormat.class);
110  // Defining output Format class which is
111  // responsible to parse the dataset
112  // into a key value pair
113  job.setOutputFormatClass(TextOutputFormat.class);
114
115  job.setOutputKeyClass(Text.class);
116  job.setOutputValueClass(IntWritable.class);
117  FileInputFormat.addInputPath(job, new Path(args[0]));
118  FileOutputFormat.setOutputPath(job, new Path(args[1]));
119  System.exit(job.waitForCompletion(true) ? 0 : 1);
120  }
121  }
```

Name: Pooja Pathak

Roll No: 14

MSc Part-I Sem-2

```
temperature1 (~) - gedit
File Edit View Search Tools Documents Help
Open Save Undo
temperature1
0067011990999992020051507004+68750+023550FM-12+038299999V0203301N00671220001CN9999999N9+00881+99999999999
0043011990999992020051512004+68750+023550FM-12+038299999V0203201N00671220001CN9999999N9+00991+99999999999
0043011990999992021051518004+68750+023550FM-12+038299999V0203201N00261220001CN9999999N9-01001+99999999999
0043012650999992021032412004+62300+010750FM-12+048599999V0202701N00461220001CN0500001N9+01111+99999999999
0043012650999992022032418004+62300+010750FM-12+048599999V0202701N00461220001CN0500001N9+01011+99999999999
0043012650999992022032418004+62300+010750FM-12+048599999V0202701N00461220001CN0500001N9+01221+99999999999
```

```
cloudera@quick
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 8 items
drwxr-xr-x - hbase supergroup 0 2022-02-07 21:29 /hbase
drwxr-xr-x - cloudera supergroup 0 2022-02-17 20:08 /inputcount
drwxr-xr-x - cloudera supergroup 0 2022-02-14 20:58 /rjc2122
drwxr-xr-x - cloudera supergroup 0 2022-02-14 21:05 /rjs2122
drwxr-xr-x - solr solr 0 2015-06-09 03:38 /solr
drwxrwxrwx - hdfs supergroup 0 2022-02-07 21:31 /tmp
drwxr-xr-x - hdfs supergroup 0 2015-06-09 03:38 /user
drwxr-xr-x - hdfs supergroup 0 2015-06-09 03:36 /var

drwxr-xr-x - nats supergroup 0 2015-06-09 03:36 /var
[cloudera@quickstart ~]$ hdfs dfs -mkdir /tempip
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 9 items
drwxr-xr-x - hbase supergroup 0 2022-02-07 21:29 /hbase
drwxr-xr-x - cloudera supergroup 0 2022-02-17 20:08 /inputcount
drwxr-xr-x - cloudera supergroup 0 2022-02-14 20:58 /rjc2122
drwxr-xr-x - cloudera supergroup 0 2022-02-14 21:05 /rjs2122
drwxr-xr-x - solr solr 0 2015-06-09 03:38 /solr
drwxr-xr-x - cloudera supergroup 0 2022-02-21 19:16 /tempip
drwxrwxrwx - hdfs supergroup 0 2022-02-07 21:31 /tmp
drwxr-xr-x - hdfs supergroup 0 2015-06-09 03:38 /user
drwxr-xr-x - hdfs supergroup 0 2015-06-09 03:36 /var
```

Name: Pooja Pathak

Roll No: 14

MSc Part-I Sem-2

```
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/temperature /tempip/
[cloudera@quickstart ~]$ hdfs dfs -ls /tempip
Found 1 items
-rw-r--r-- 1 cloudera supergroup 530 2022-02-21 19:19 /tempip/temperature
[cloudera@quickstart ~]$ hdfs dfs -cat /tempip/temperature
0067011990999991950051507004+68750+023550FM-12+038299999V0203301N00671220001CN99
99999N9+00001+99999999999
0043011990999991950051512004+68750+023550FM-12+038299999V0203201N00671220001CN99
99999N9+00221+99999999999
0043011990999991950051518004+68750+023550FM-12+038299999V0203201N00671220001CN99
99999N9-00111+99999999999
0043012650999991949032412004+62300+010750FM-12+048599999V0202701N00461220001CN05
00001N9+01111+99999999999
0043012650999991949032418004+62300+010750FM-12+048599999V0202701N00461220001CN05
00001N9+000781+99999999999
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/MaxTemp.jar MaxTemp /tempip/t
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/MaxTemp.jar MaxTemp /tempip/t
emperature /tempip1
22/02/21 19:25:47 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0
:8032
22/02/21 19:25:48 WARN mapreduce.JobSubmitter: Hadoop command-line option parsin
g not performed. Implement the Tool interface and execute your application with
ToolRunner to remedy this.
22/02/21 19:25:48 INFO input.FileInputFormat: Total input paths to process : 1
22/02/21 19:25:48 INFO mapreduce.JobSubmitter: number of splits:1
22/02/21 19:25:49 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_16
44298092376_0001
22/02/21 19:25:49 INFO impl.YarnClientImpl: Submitted application application_16
44298092376_0001
22/02/21 19:25:49 INFO mapreduce.Job: The url to track the job: http://quickstar
t.cloudera:8088/proxy/application_1644298092376_0001/
22/02/21 19:25:49 INFO mapreduce.Job: Running job: job_1644298092376_0001
22/02/21 19:26:04 INFO mapreduce.Job: Job job_1644298092376_0001 running in uber
mode : false
22/02/21 19:26:04 INFO mapreduce.Job: map 0% reduce 0%
22/02/21 19:26:15 INFO mapreduce.Job: map 100% reduce 0%
22/02/21 19:26:26 INFO mapreduce.Job: map 100% reduce 100%
22/02/21 19:26:26 INFO mapreduce.Job: Job job_1644298092376_0001 completed succe
ssfully
```

Name: Pooja Pathak

Roll No: 14

MSc Part-I Sem-2

22/02/21 19:26:26 INFO mapreduce.Job: Job job_1644298092376_0001 completed successfully

22/02/21 19:26:26 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=61
FILE: Number of bytes written=221207
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=645
HDFS: Number of bytes written=17
HDFS: Number of read operations=6
HDFS: Number of large read operations=0
HDFS: Number of write operations=2

Job Counters

Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=9746
Total time spent by all reduces in occupied slots (ms)=7688
Total time spent by all map tasks (ms)=9746
Total time spent by all reduce tasks (ms)=7688
Total vcore-seconds taken by all map tasks=9746
Total vcore-seconds taken by all reduce tasks=7688
Total megabyte-seconds taken by all map tasks=9979904
Total megabyte-seconds taken by all reduce tasks=7872512

Map-Reduce Framework

Map input records=5
Map output records=5
Map output bytes=45
Map output materialized bytes=61
Input split bytes=115
Combine input records=0
Combine output records=0
Reduce input groups=2
Reduce shuffle bytes=61
Reduce input records=5
Reduce output records=2
Spilled Records=10
Shuffled Maps =1

Name: Pooja Pathak

Roll No: 14

MSc Part-I Sem-2

```
File Edit View Search Terminal Help
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=204
CPU time spent (ms)=1480
Physical memory (bytes) snapshot=328798208
Virtual memory (bytes) snapshot=3007225856
Total committed heap usage (bytes)=226365440
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=530
File Output Format Counters
Bytes Written=17
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 10 items
drwxr-xr-x - hbase supergroup 0 2022-02-07 21:29 /hbase
drwxr-xr-x - cloudera supergroup 0 2022-02-17 20:08 /inputcount
drwxr-xr-x - cloudera supergroup 0 2022-02-14 20:58 /rjc2122
drwxr-xr-x - cloudera supergroup 0 2022-02-14 21:05 /rjs2122
drwxr-xr-x - solr solr 0 2015-06-09 03:38 /solr
drwxr-xr-x - cloudera supergroup 0 2022-02-21 19:19 /tempip
drwxr-xr-x - cloudera supergroup 0 2022-02-21 19:26 /tempop1
drwxrwxrwx - hdfs supergroup 0 2022-02-07 21:31 /tmp
drwxr-xr-x - hdfs supergroup 0 2015-06-09 03:38 /user
drwxr-xr-x - hdfs supergroup 0 2015-06-09 03:36 /var
[cloudera@quickstart ~]$ hdfs dfs -ls /tempop1
Found 2 items
-rw-r--r-- 1 cloudera supergroup 0 2022-02-21 19:26 /tempop1/_SUCCESS
-rw-r--r-- 1 cloudera supergroup 17 2022-02-21 19:26 /tempop1/part-r-00000
[cloudera@quickstart ~]$ hdfs dfs -cat /tempop1/part-r-00000
1949 111
1950 22
[cloudera@quickstart ~]$
```

Name: Pooja Pathak

Roll No: 14

MSc Part-I Sem-2