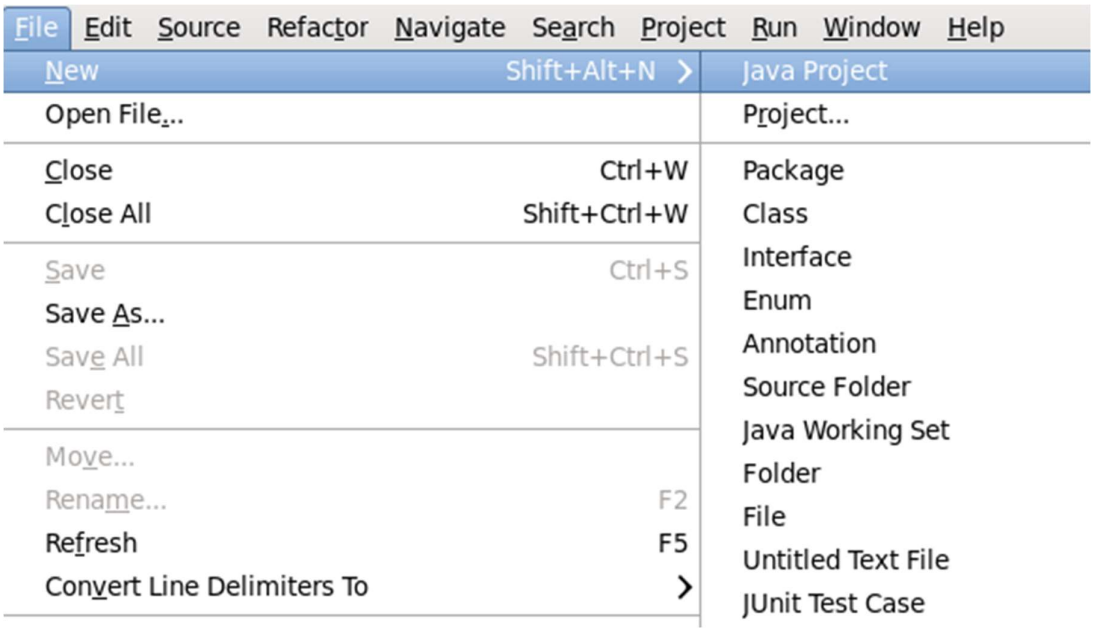
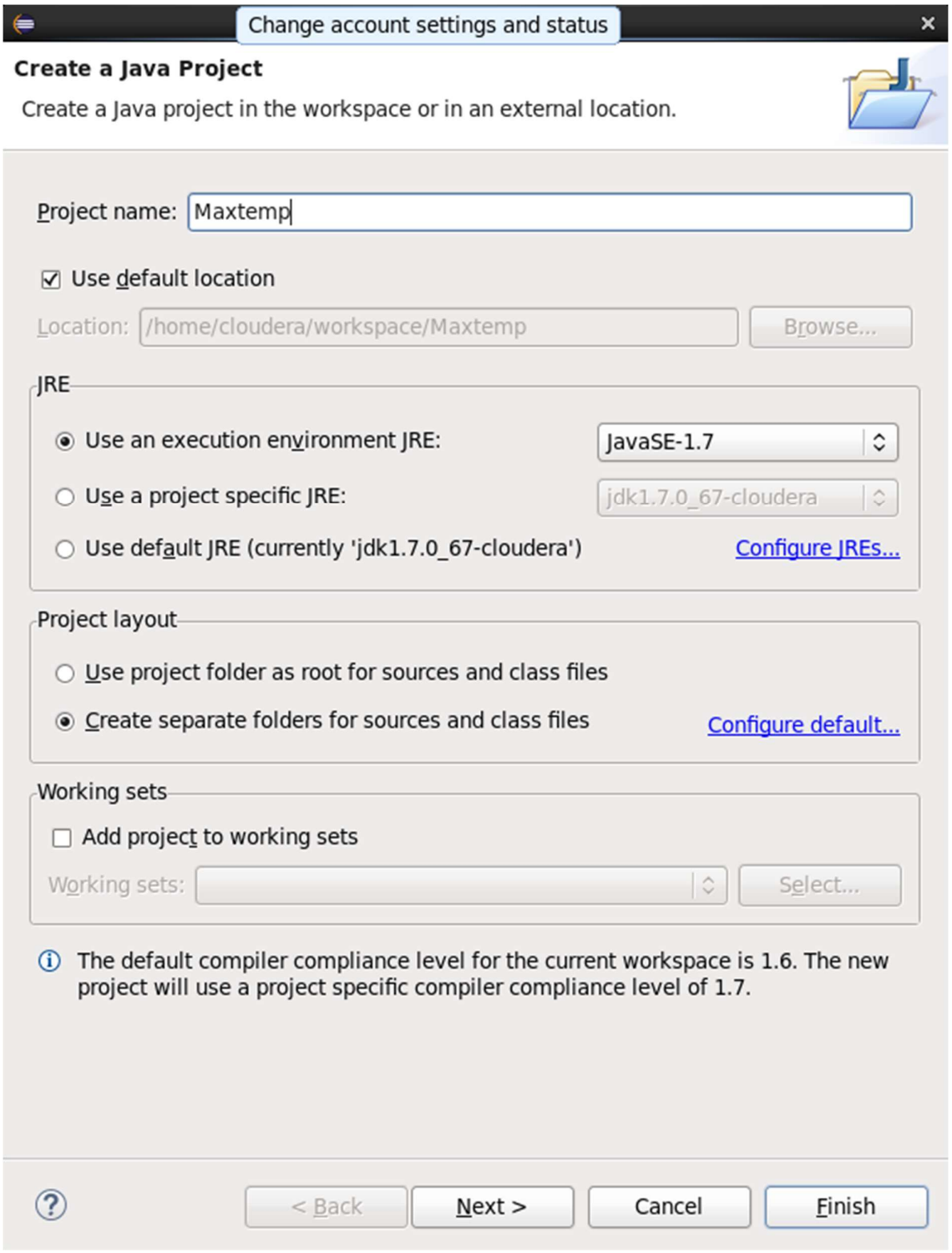
Name: Pooja Pathak
Roll No: 14
MSc Part-I Sem-2

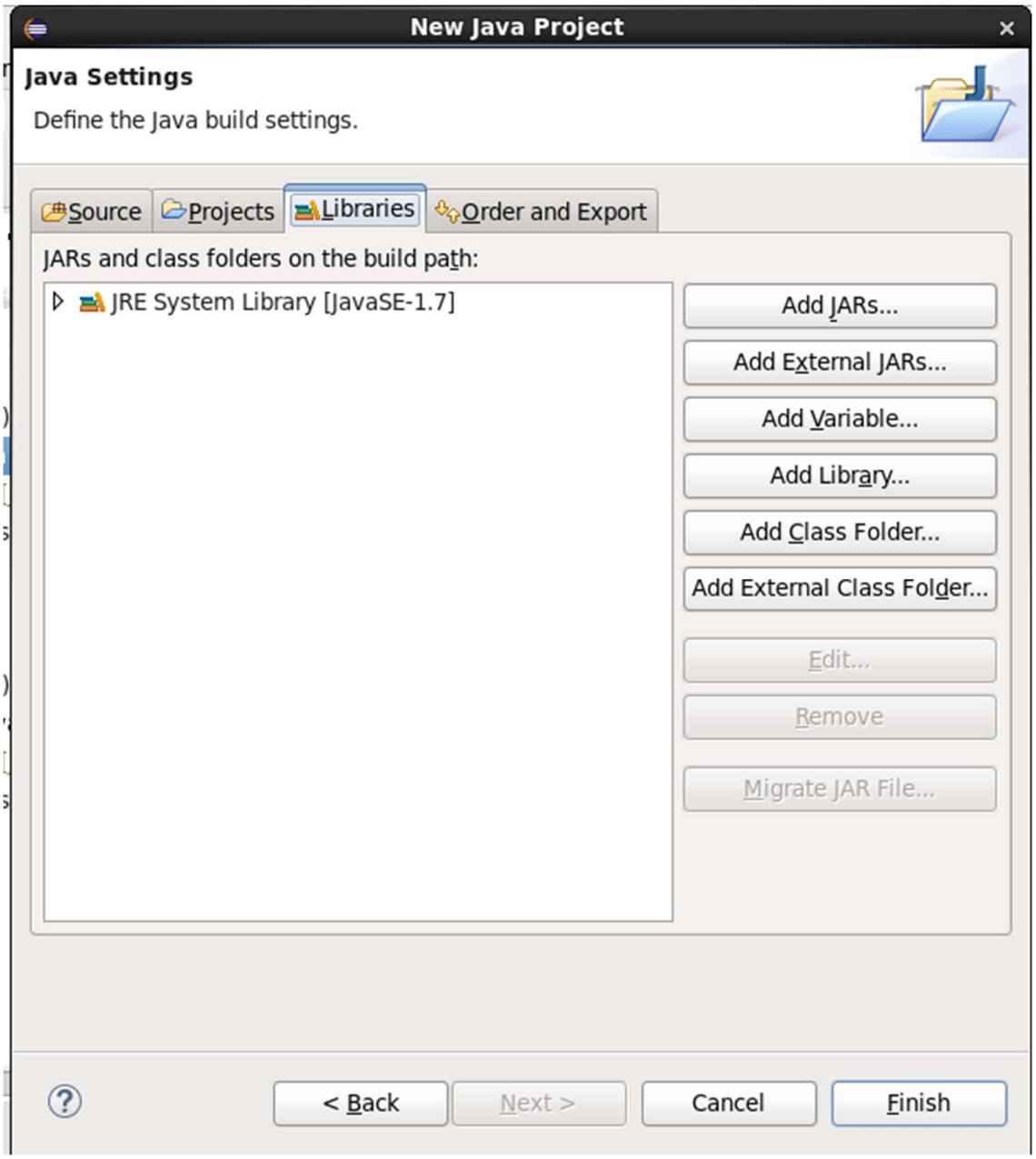
Subject: Big Data Technology
Practical 4: Max Temp



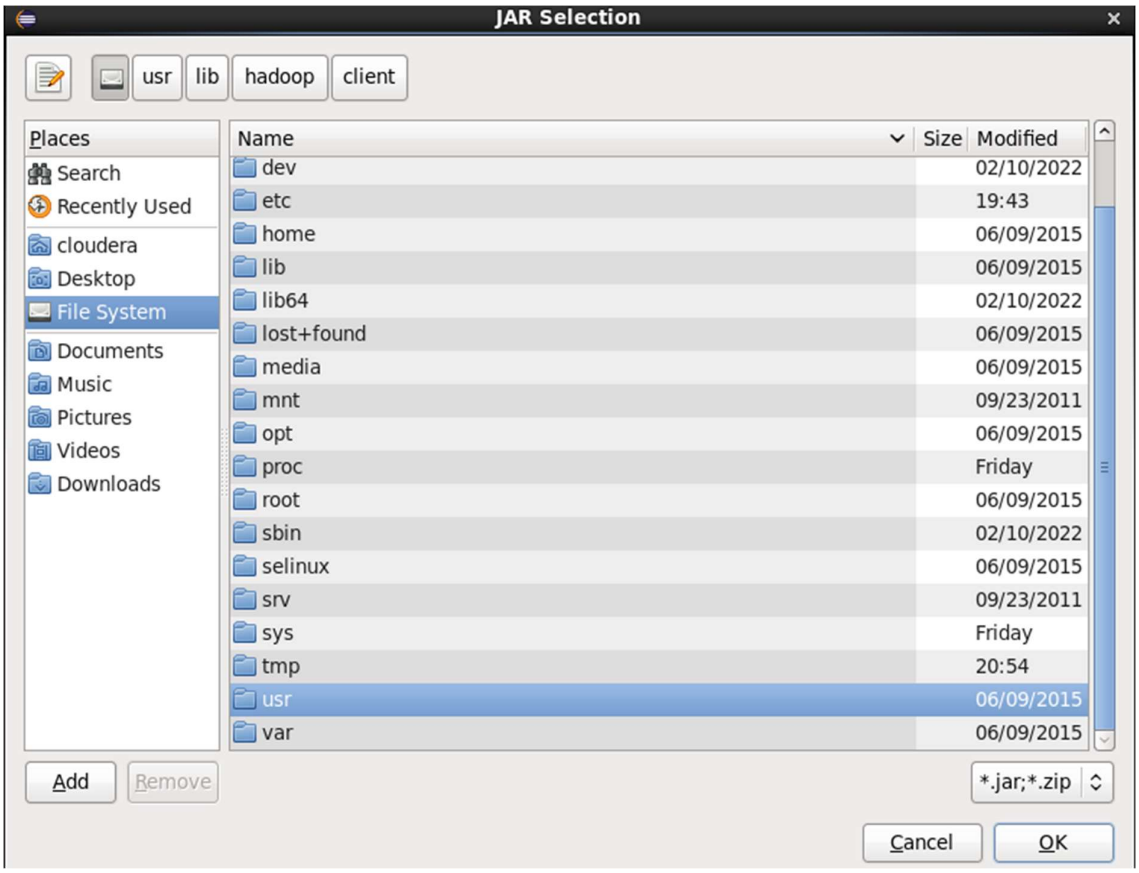
Name: Pooja Pathak
Roll No: 14
MSc Part-I Sem-2



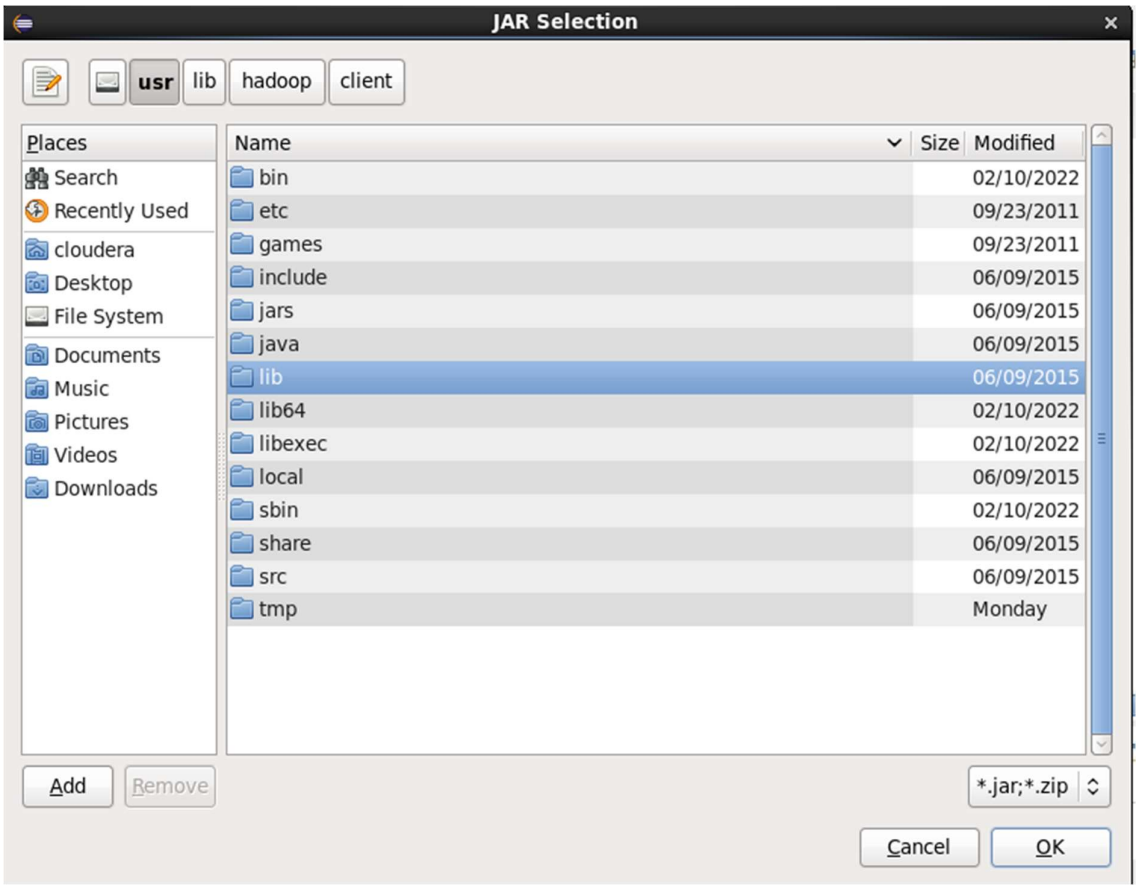
Name: Pooja Pathak
Roll No: 14
MSc Part-I Sem-2



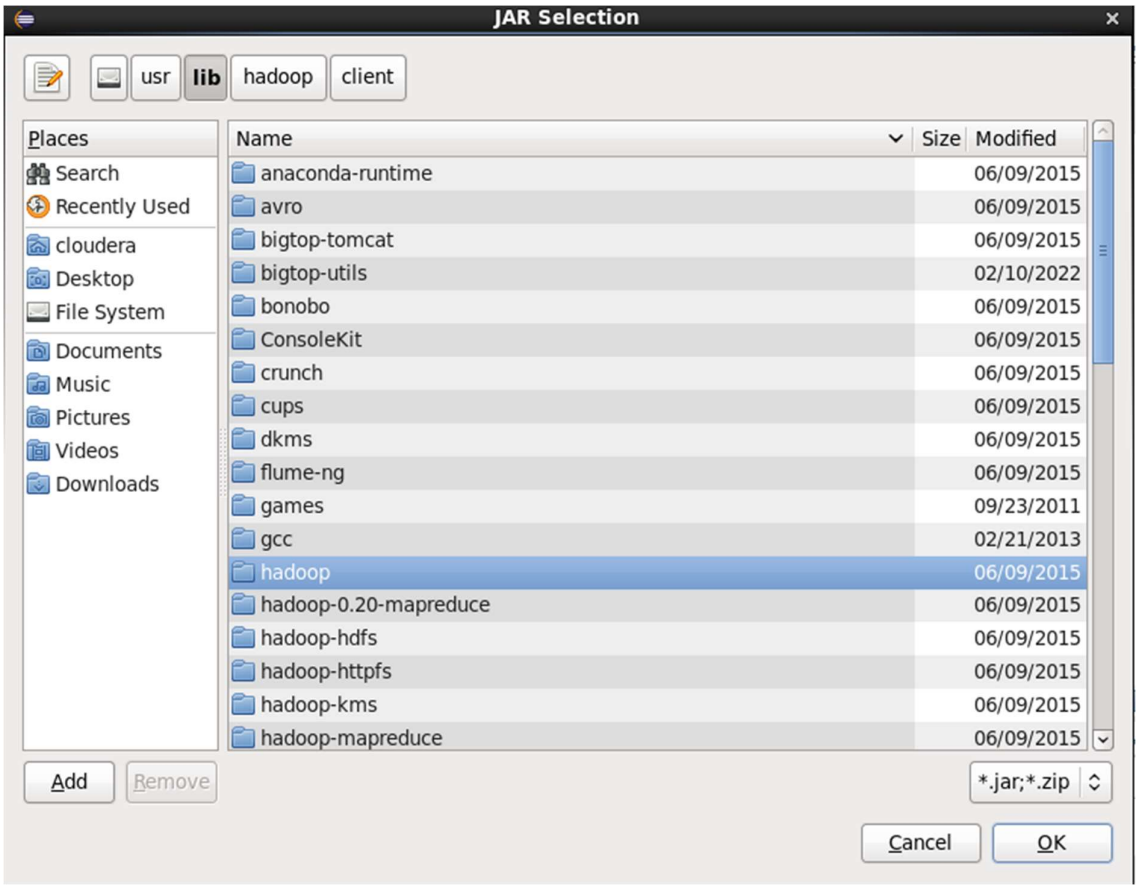
Name: Pooja Pathak
Roll No: 14
MSc Part-I Sem-2



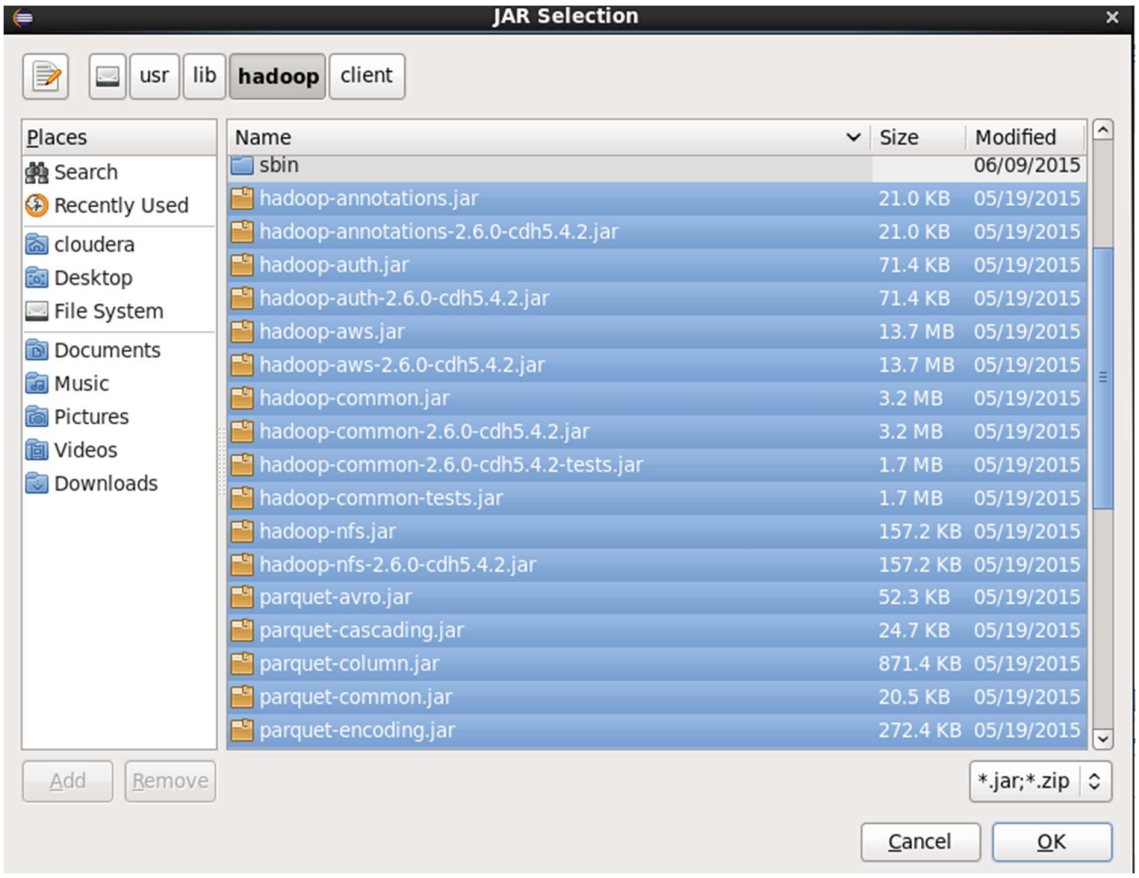
Name: Pooja Pathak
Roll No: 14
MSc Part-I Sem-2



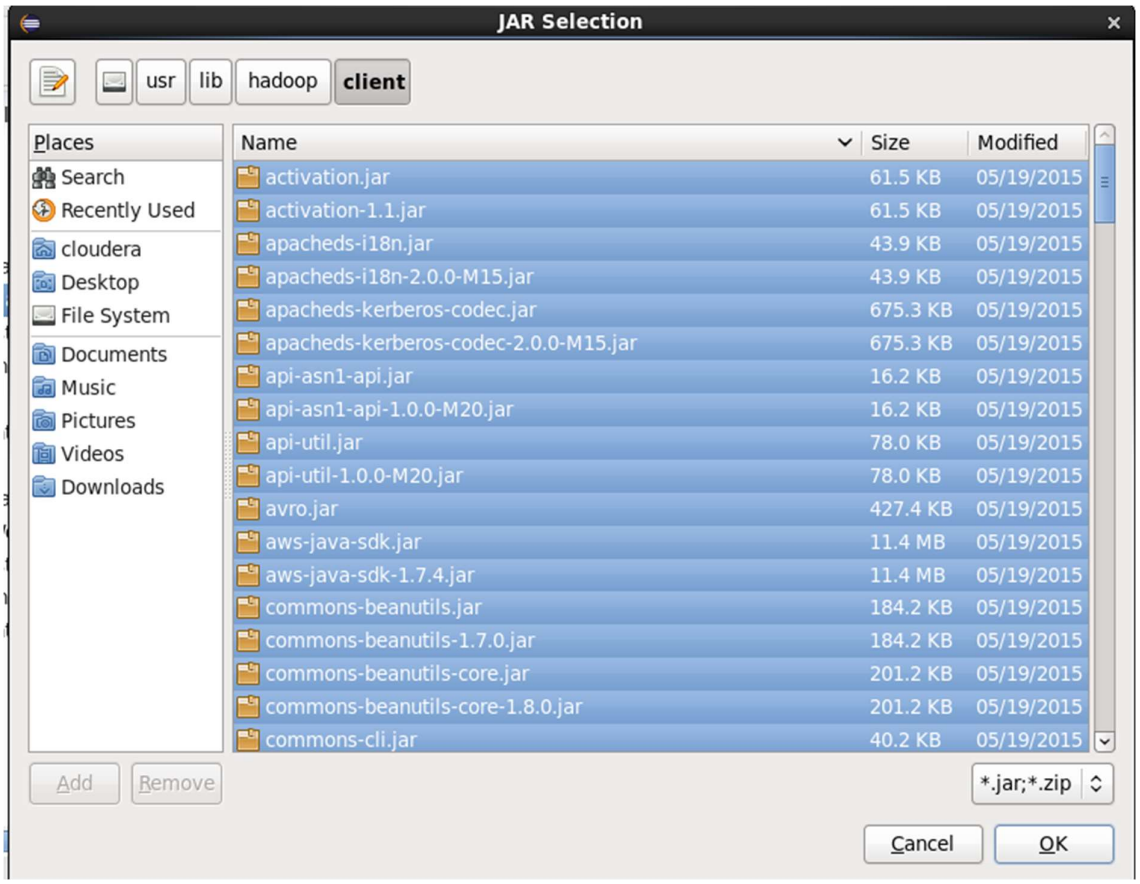
Name: Pooja Pathak
Roll No: 14
MSc Part-I Sem-2



Name: Pooja Pathak
Roll No: 14
MSc Part-I Sem-2



Name: Pooja Pathak
Roll No: 14
MSc Part-I Sem-2



Name: Pooja Pathak
Roll No: 14
MSc Part-I Sem-2

StubDriver.java | WordCount.java | MaxTemp.java

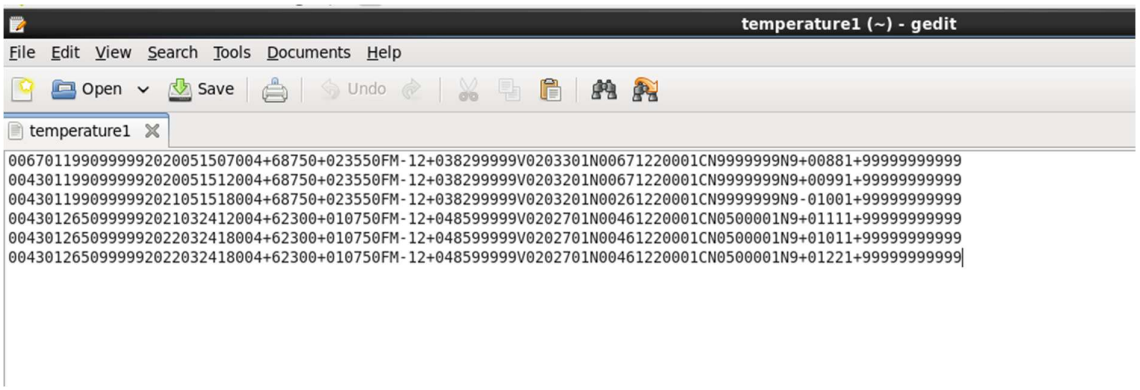
```
1  import java.io.IOException;
18 //import MaximumTemp.MaxTemperatureMapper;
19 //import MaxTemp.MaxTemperatureReducer;
20
21 public class MaxTemp {
22     // Mapper
23     /*MaxTemperatureMapper class is static
24     * and extends Mapper abstract class
25     * having four Hadoop generics type
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
```

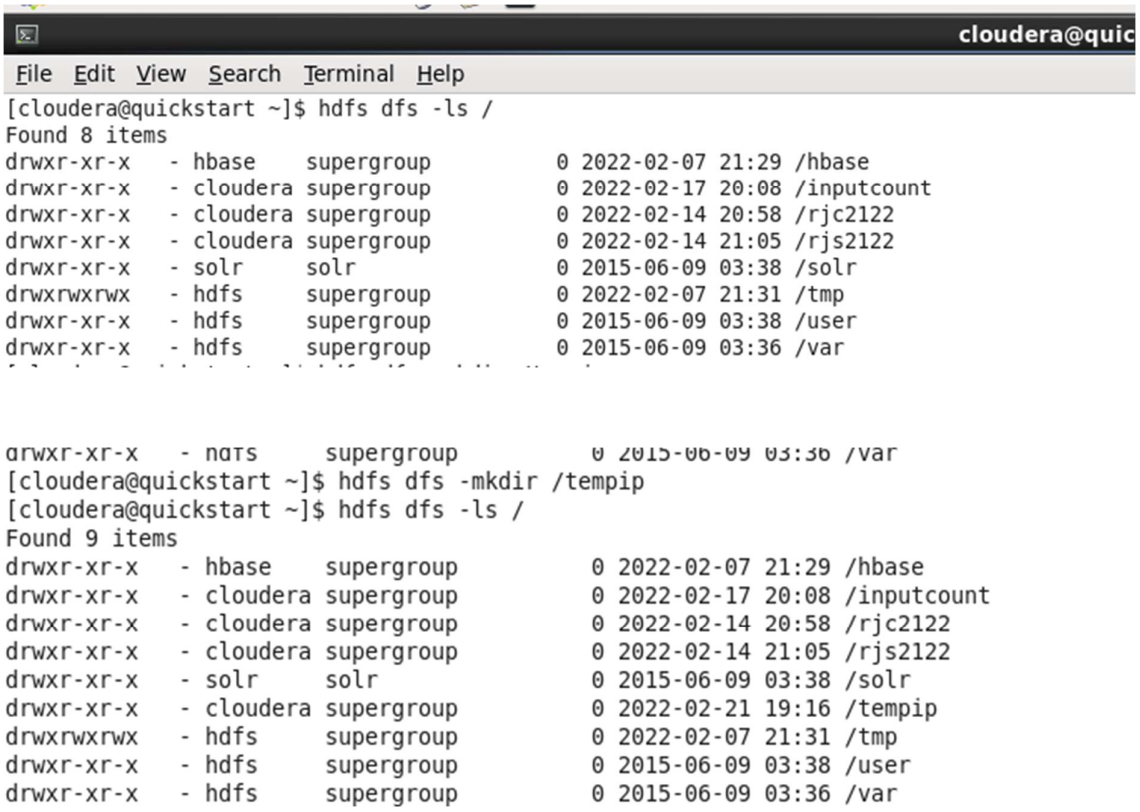
Name: Pooja Pathak
Roll No: 14
MSc Part-I Sem-2

```
    * @method reduce
    * This method takes the input as key and
    * list of values pair from the mapper,
    * it does aggregation based on keys and
    * produces the final context.
    */
    public void reduce(Text key,Iterable<IntWritable> values, Context context)
    throws IOException, InterruptedException {
    int maxvalue= Integer.MIN_VALUE;
    for (IntWritable value : values) {
    maxvalue=Math.max(maxvalue, value.get());
    }
    context.write(key, new IntWritable(maxvalue));
    }
    }

    /**
    * @method main
    * This method is used for setting
    * all the configuration properties.
    * It acts as a driver for map-reduce
    * code.
    */
    public static void main(String[] args) throws Exception {
    // reads the default configuration of the
    // cluster from the configuration XML files
    Configuration conf = new Configuration();
    // Initializing the job with the
    // default configuration of the cluster
    // Job job = new Job(conf, "weather example");
    Job job = Job.getInstance(conf, "weather example");
    // Assigning the driver class name
    job.setJarByClass(MaxTemp.class);
    // Key type coming out of mapper
    // job.setMapOutputKeyClass(Text.class);
    // value type coming out of mapper
    // job.setMapOutputValueClass(Text.class);
    // Defining the mapper class name
    job.setMapperClass(MaxTemperatureMapper.class);
    // Defining the reducer class name
    job.setReducerClass(MaxTemperatureReducer.class);
    // Defining input Format class which is
    // responsible to parse the dataset
    // into a key value pair
    job.setInputFormatClass(TextInputFormat.class);
    // Defining output Format class which is
```

```
job.setInputFormatClass(TextInputFormat.class);
// Defining output Format class which is
// responsible to parse the dataset
// into a key value pair
job.setOutputFormatClass(TextOutputFormat.class);

job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class);
FileInputFormat.addInputPath(job, new Path(args[0]));
FileOutputFormat.setOutputPath(job, new Path(args[1]));
System.exit(job.waitForCompletion(true) ? 0 : 1);
}
}
```

Name: Pooja Pathak
Roll No: 14
MSc Part-I Sem-2





Name: Pooja Pathak
Roll No: 14
MSc Part-I Sem-2

```
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/temperature /tempi
p/
[cloudera@quickstart ~]$ hdfs dfs -ls /tempip
Found 1 items
-rw-r--r--   1 cloudera supergroup        530 2022-02-21 19:19 /tempip/temperatu
re
[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+038299999V0203201N00261220001CN99
99999N9-00111+99999999999
0043012650999991949032412004+62300+010750FM-12+048599999V0202701N00461220001CN05
00001N9+01111+99999999999
0043012650999991949032418004+62300+010750FM-12+048599999V0202701N00461220001CN05
00001N9+00781+99999999999
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/MaxTemp.jar MaxTemp /tempip/t
emperature /tempop1
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 succe
ssfully
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-0
0000
[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