



Wisdom Sprouts

We foster knowledge here

---

BIG DATA HADOOP

---

Lab Manual



## Executing MapReduce Word Count Program

*Step 1: Create an input folder in HDFS for your data which needs to be processed*

```
hadoop fs -mkdir /user/wisdom/wordcountinput
```

*Expected Output*

Nothing

*Step 2: Put wordcount.txt file from the Linux file system into the newly created directory*

```
hadoop fs -put /home/wisdom/Documents/wordcount.txt wordcountinput
```

*Expected Output*

Nothing

*Step 3: Verify the contents of the newly created directory in Hadoop Hdfs*

```
hadoop fs -ls wordcountinput
```

*Expected Output*

Found 1 items

```
-rw-r--r--  1 training supergroup          293 2014-12-03 15:57
wordcountinput/wordcount.txt
```

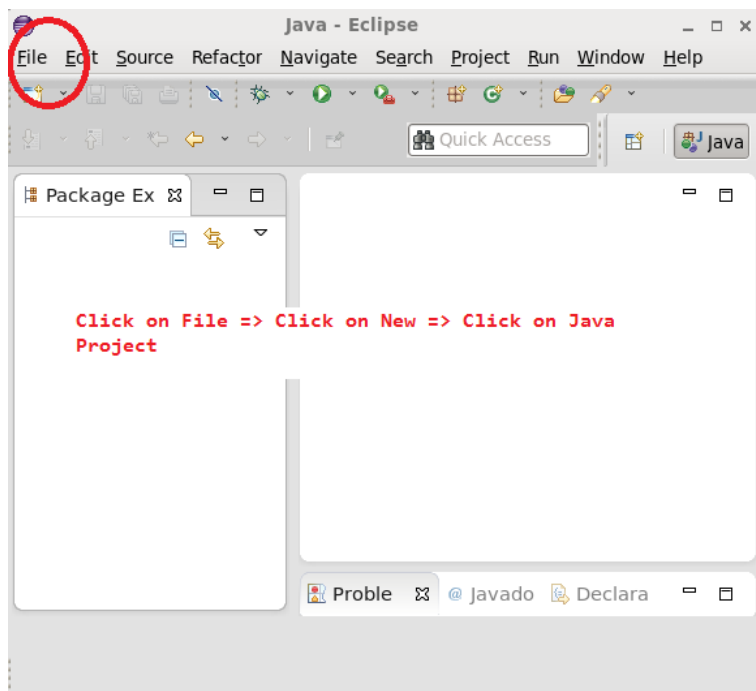
*Step 4: Copy the program's folder into the Java Workspace*

```
cp -r /home/wisdom/programs/wordcount /home/wisdom/workspace
```

*Expected Output*

Nothing

*Step 5: Follow the screenshots for creating a project in Java Eclipse and then exporting it as a Jar file*



New Java Project

Create a Java Project

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

Project name: wordcount

Enter the project Name. It should be present in the Workspace

☒ Use default location

Location: /home/wisdom/workspace/wordcount

Browse...

JRE

☒ Use an execution environment JRE:

javaSE-1.6

☐ Use a project specific JRE:

java-1.6.0-openj

☐ Use default JRE (currently 'java-1.6.0-openjdk-1.6.0.0')

Configure JREs...

Project layout

☐ Use project folder as root for sources and class files

☒ Create separate folders for sources and class files

Configure default...

Click here to complete

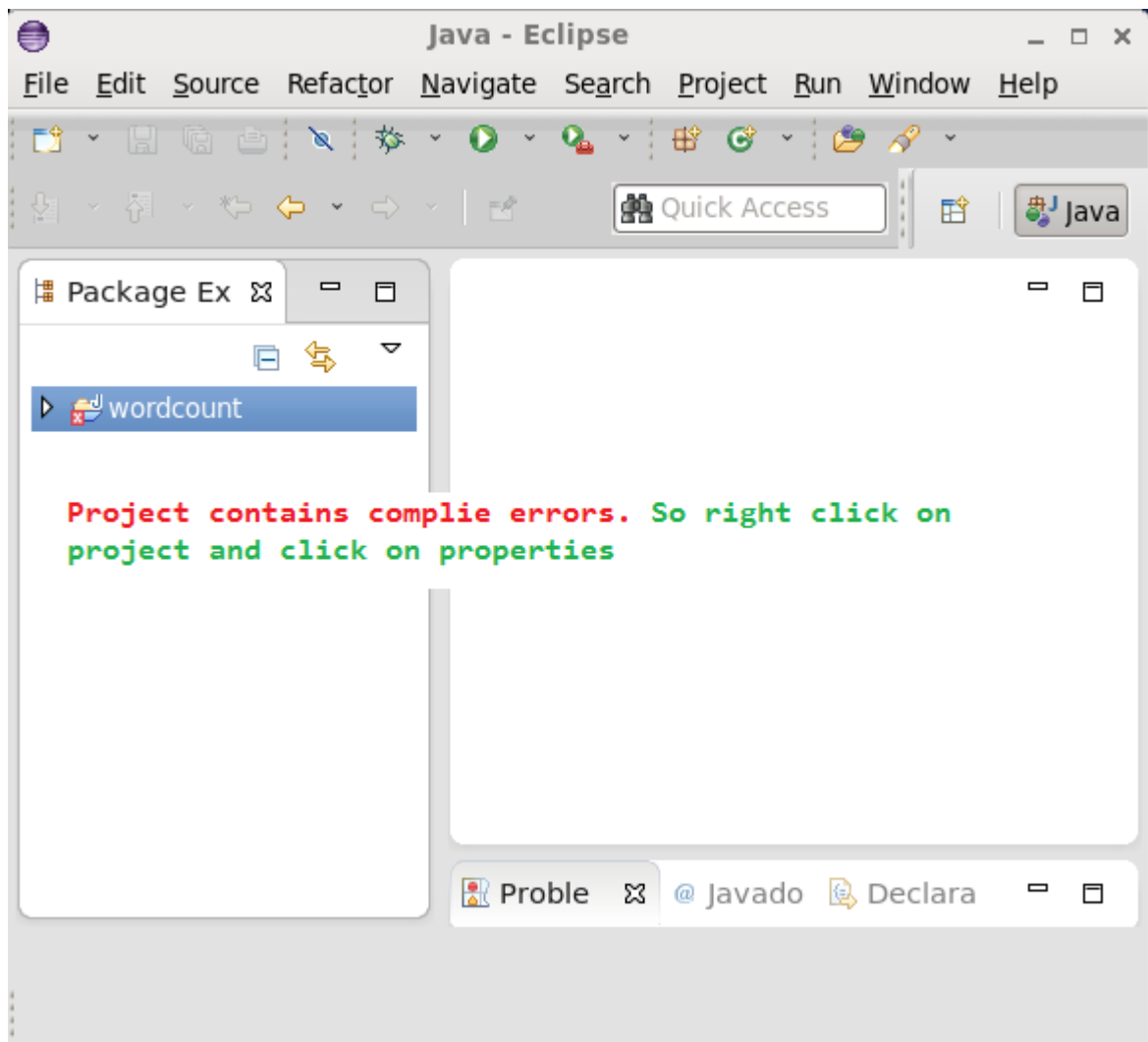
?

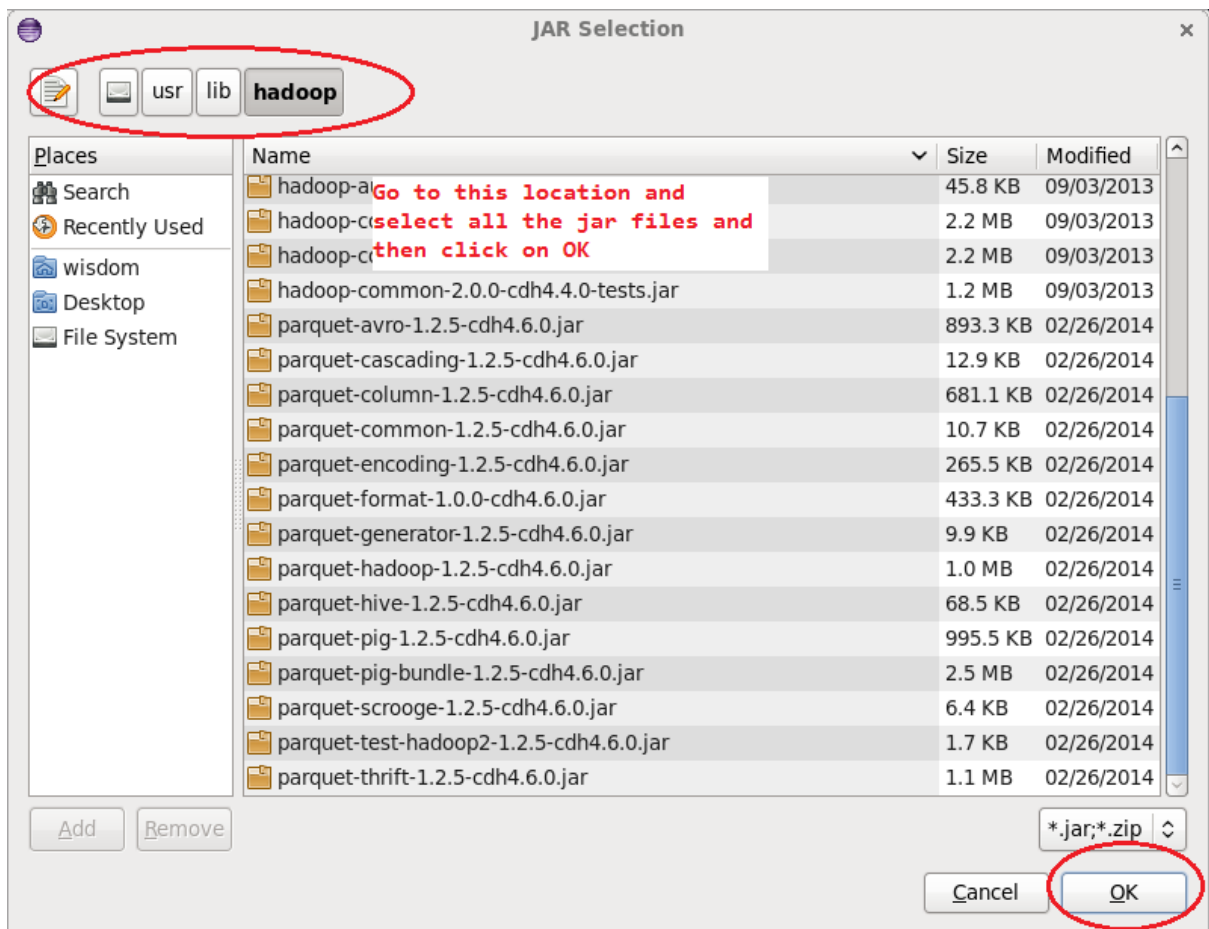
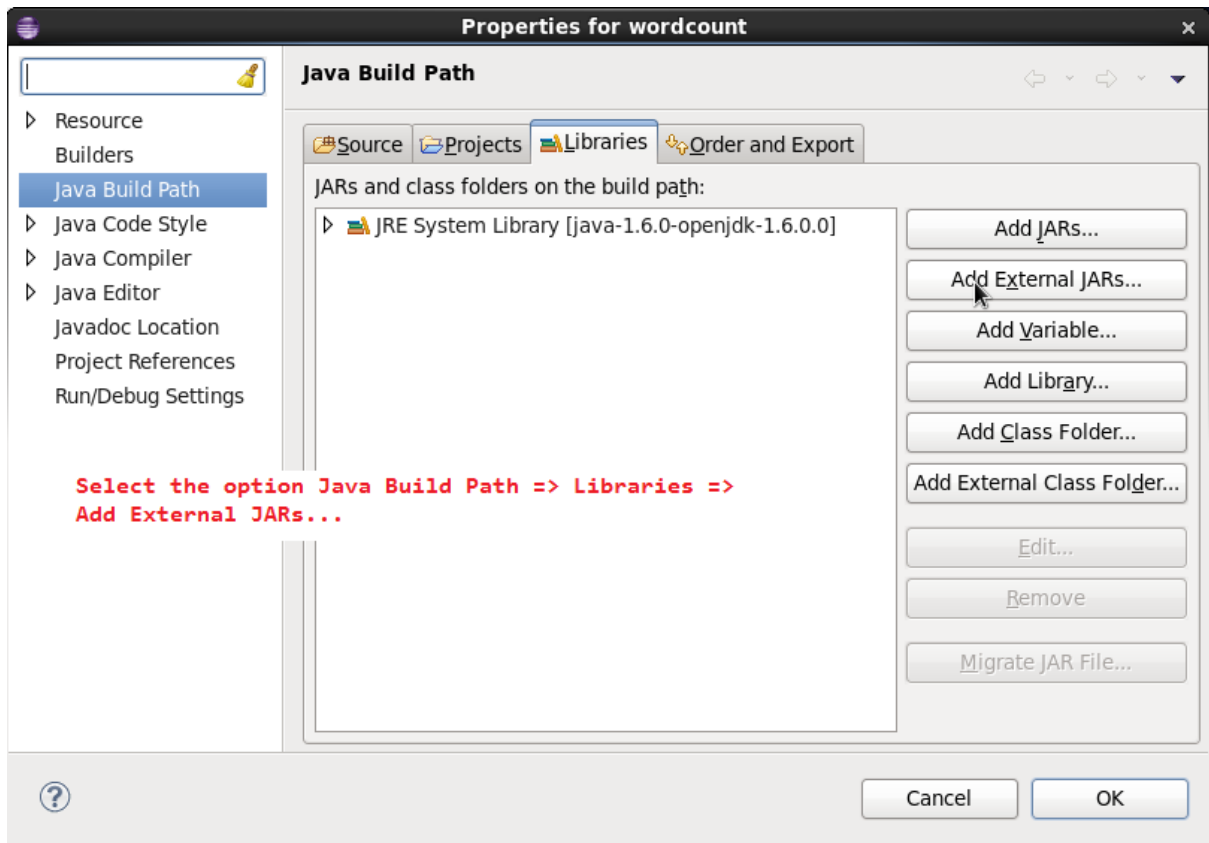
< Back

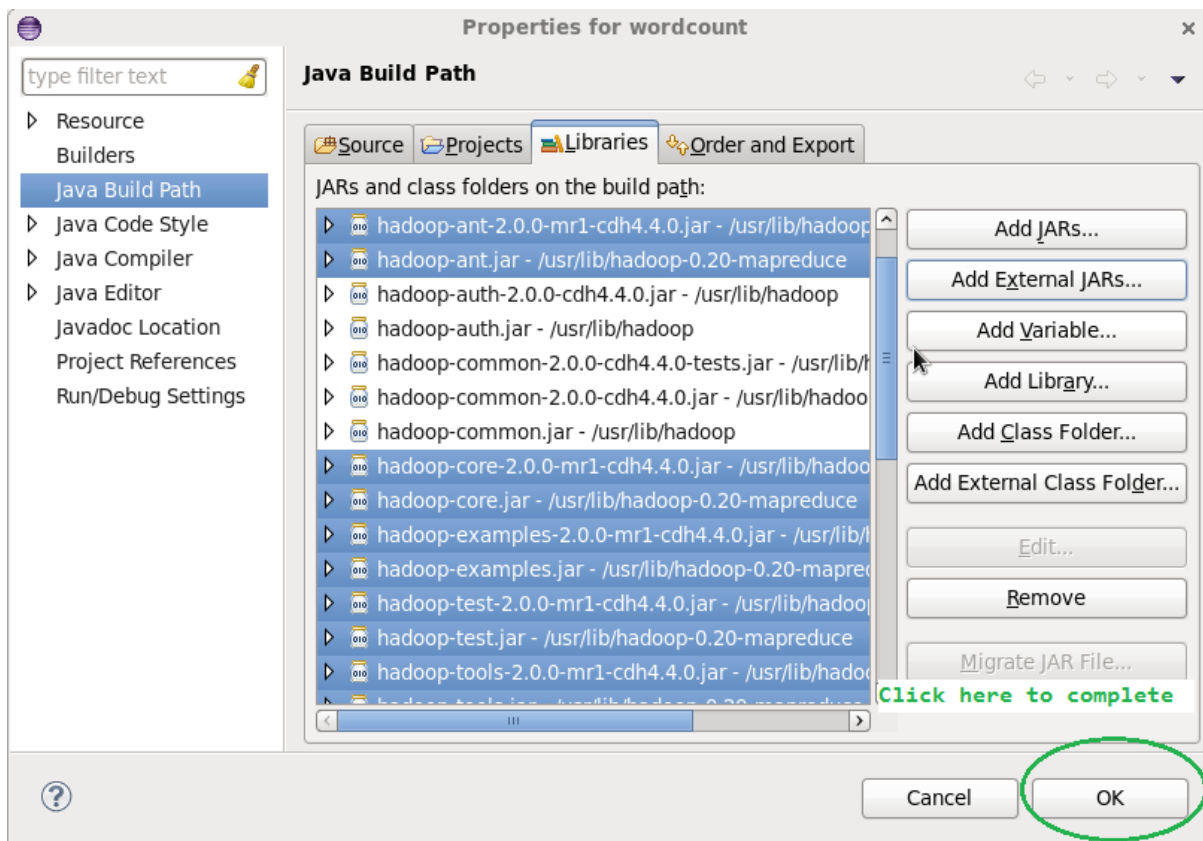
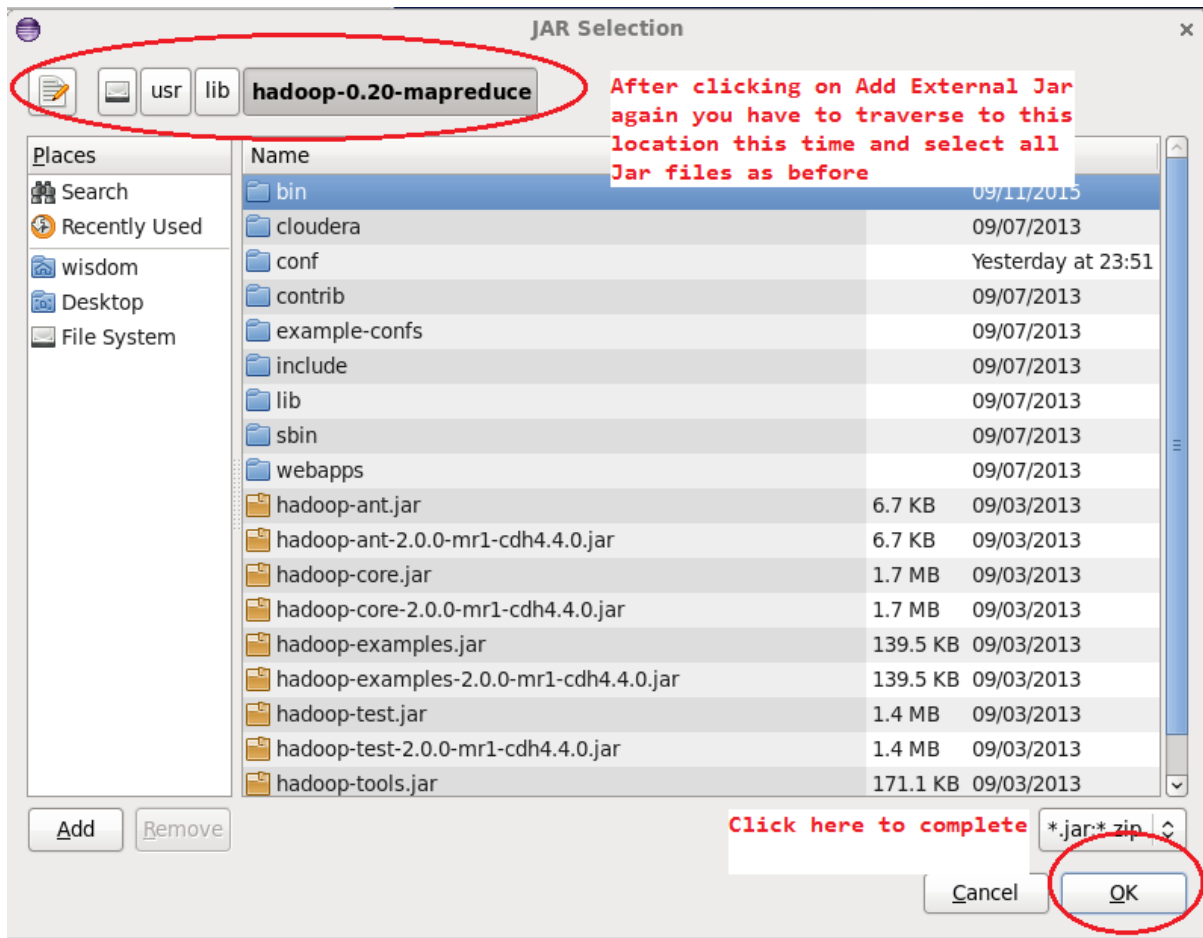
Next >

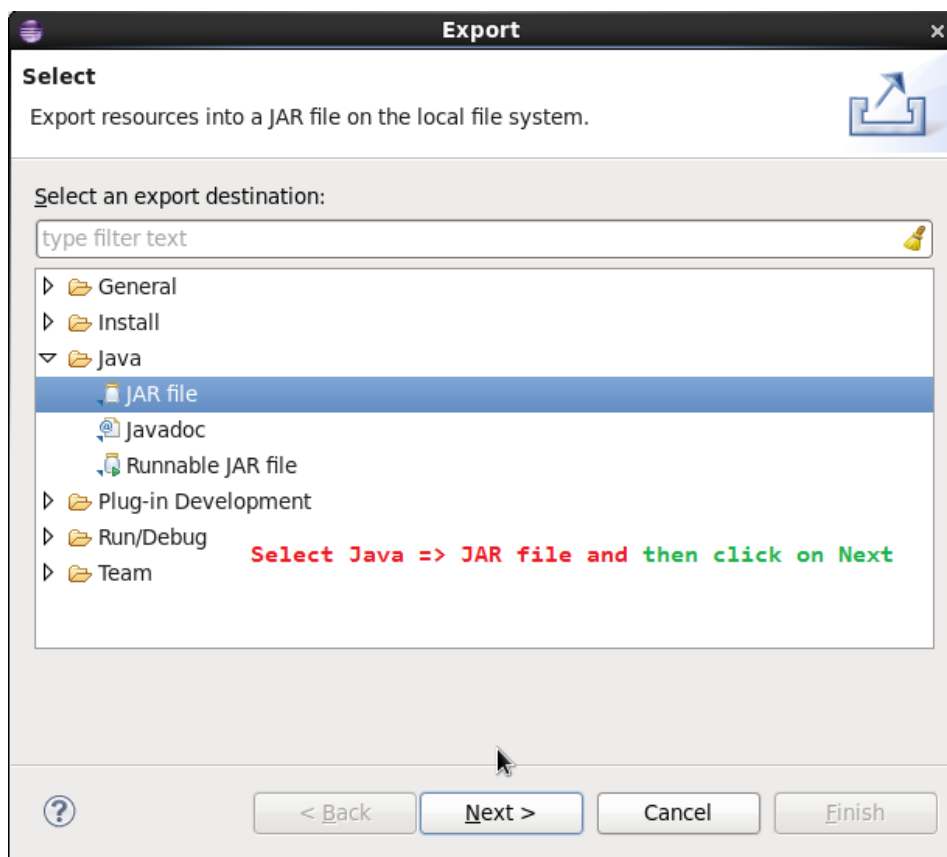
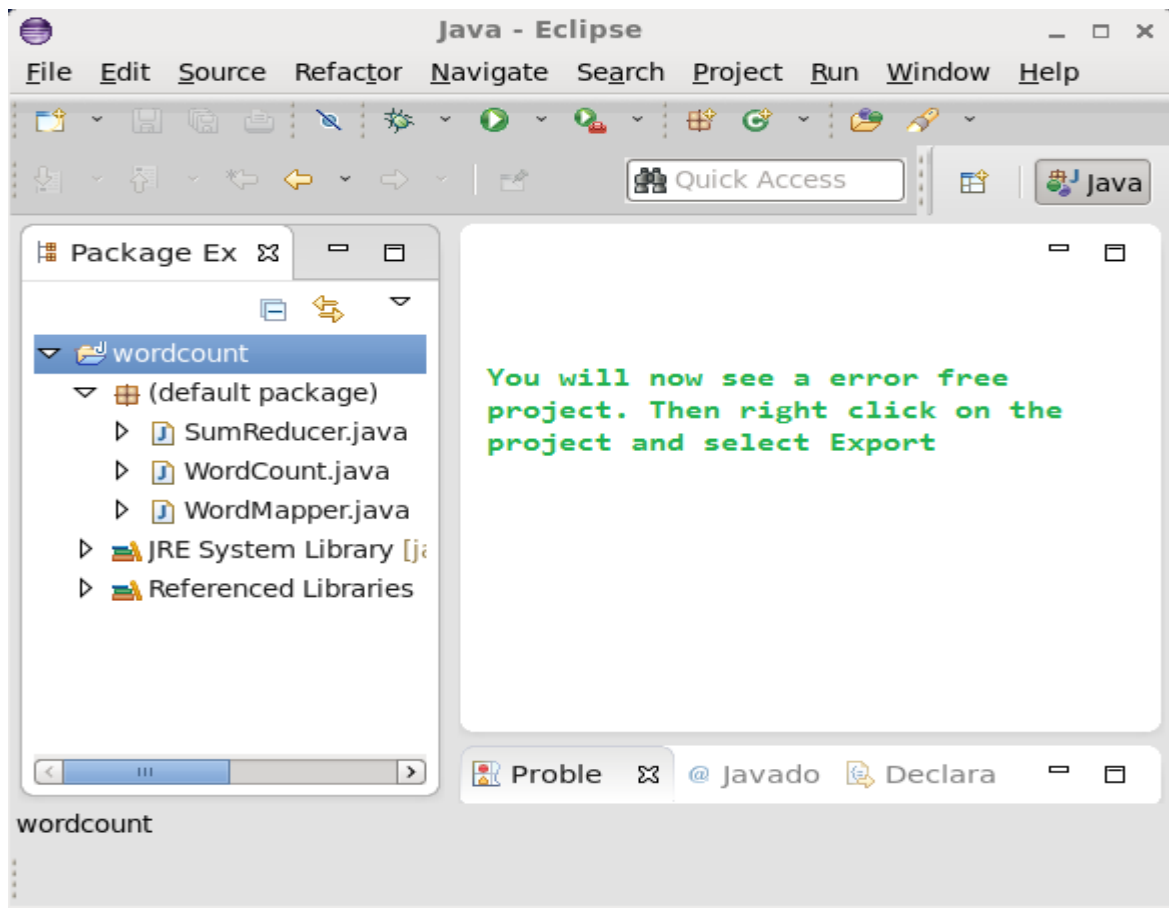
Cancel

Finish

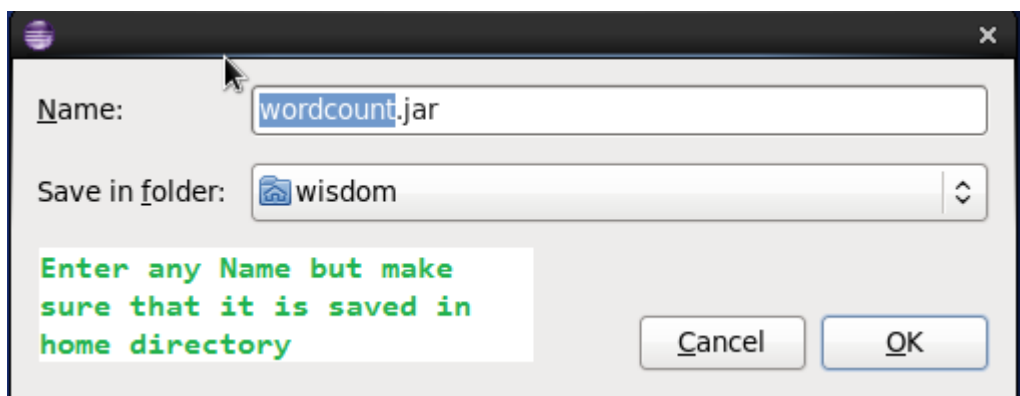
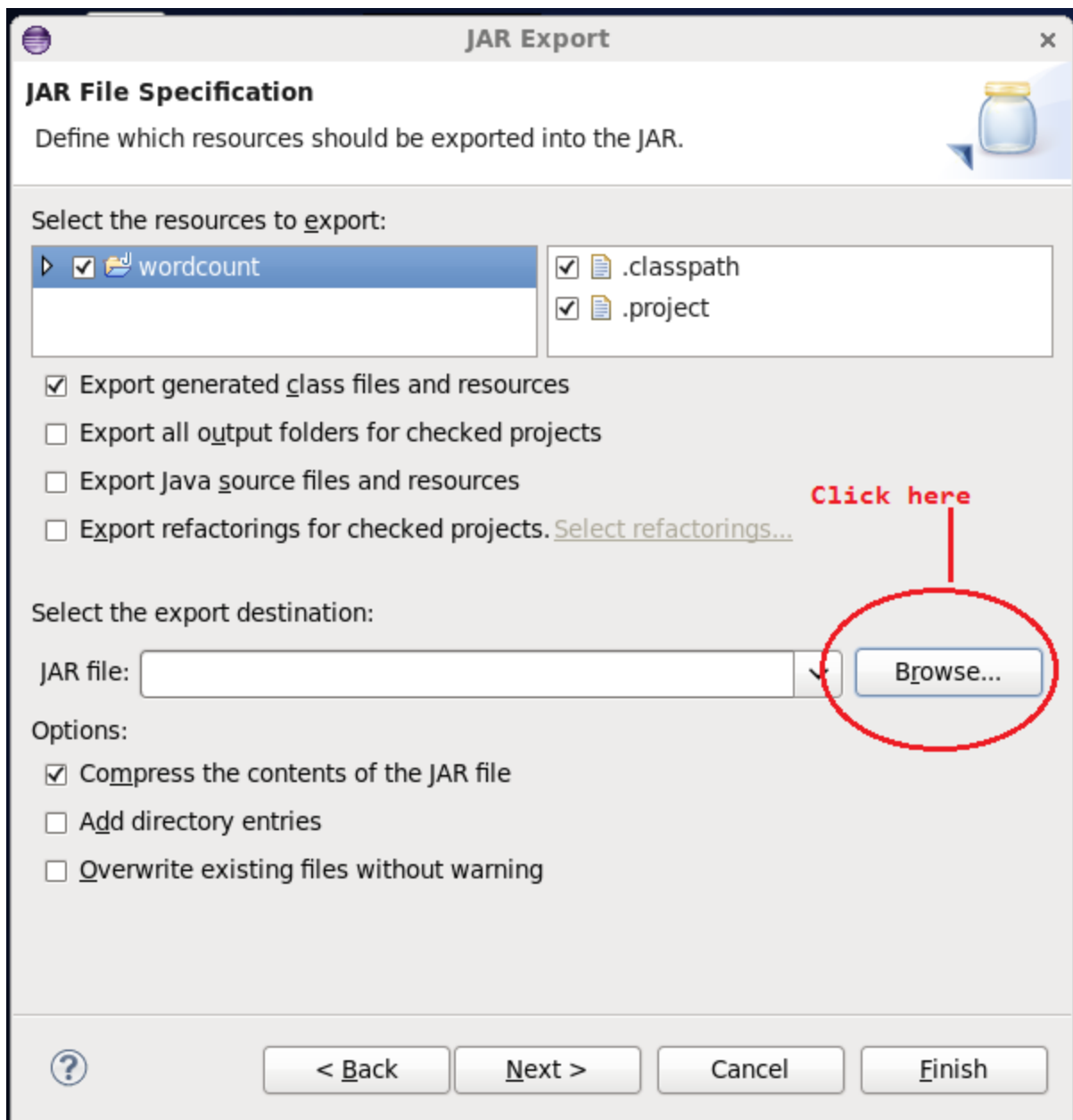


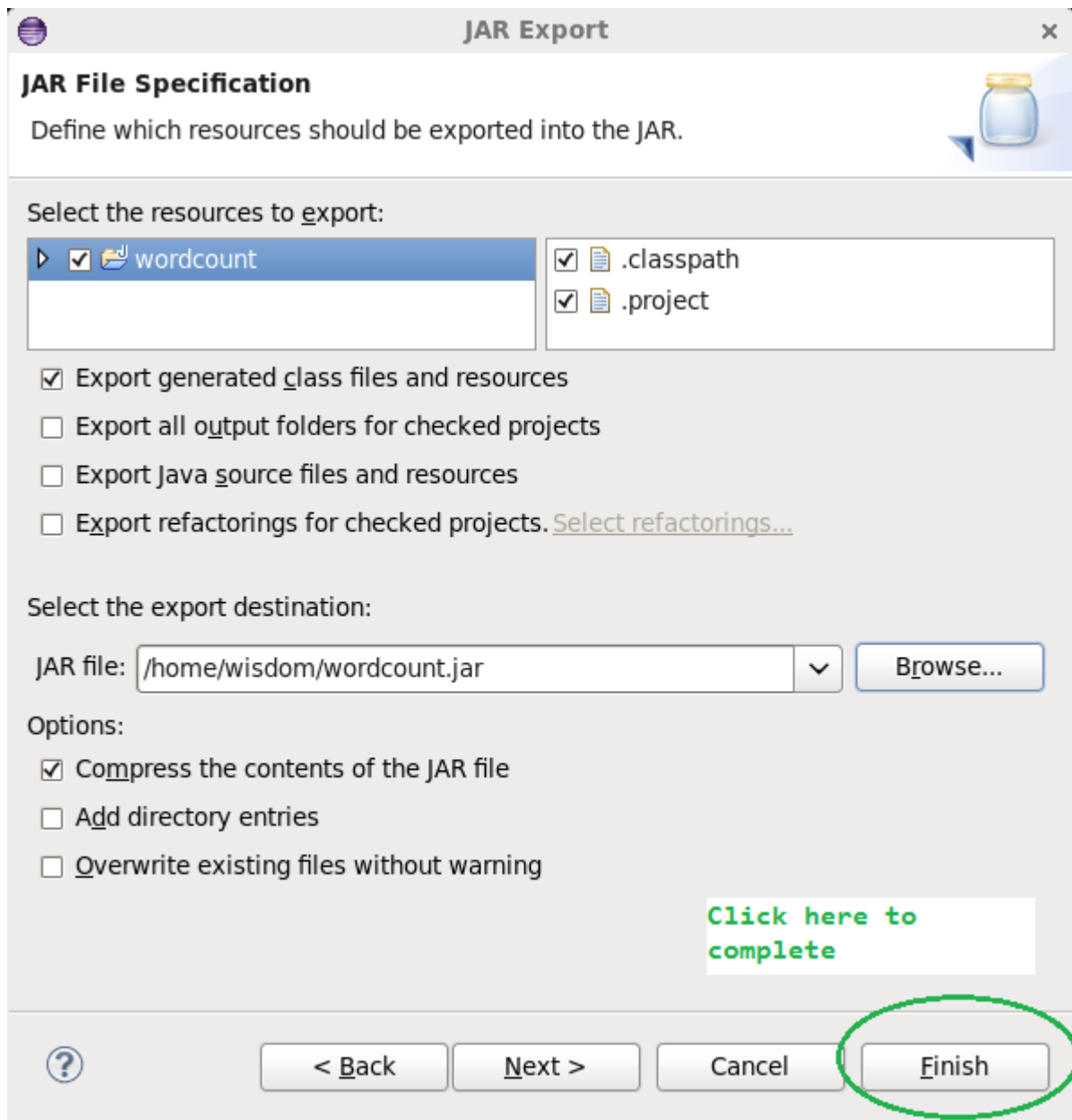












*Step 6: Run the Word Count MapReduce Program*

```
hadoop jar wordcount.jar WordCount wordcountinput wordcountoutput
```

### *Expected Output*

```
14/12/03 16:05:16 WARN mapred.JobClient: Use GenericOptionsParser
for parsing the arguments. Applications should implement Tool for
the same.
```

```
14/12/03 16:05:16 INFO mapred.FileInputFormat: Total input paths to
process : 1
```

```
14/12/03 16:05:17 INFO mapred.JobClient: Running job:
job_201412031546_0003
```

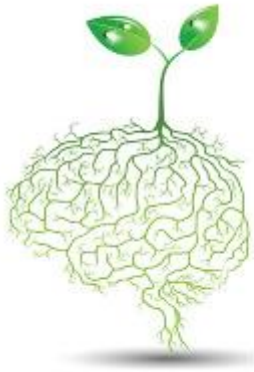
Then lots of things will happen

```
14/12/03 16:05:44 INFO mapred.JobClient:      Total committed heap
usage (bytes)=336994304
```

```
14/12/03 16:05:44 INFO mapred.JobClient:
org.apache.hadoop.mapreduce.lib.input.FileInputFormatCounter
```

```
14/12/03 16:05:44 INFO mapred.JobClient:      BYTES_READ=293
```

[www.wisdomsprouts.com](http://www.wisdomsprouts.com)



# WISDOM SPROUTS

We Foster Knowledge Here

## Contact Us

You can call us on +91-976-220-6123

Or write to us at [helpdesk@wisdomsprouts.com](mailto:helpdesk@wisdomsprouts.com)

Or Visit [www.wisdomsprouts.com](http://www.wisdomsprouts.com)