

4. Running the Wordcount Mapreduce job

Step 1: Put the data in the HDFS

```
hadoop fs -mkdir /user/training/wordcountinput  
hadoop fs -put wordcount.txt /user/training/wordcountinput
```

Step 2: Create a new project in eclipse called wordcount

1. `cp -r /home/training/exercises/wordcount /home/training/workspace/wordcount`
2. Open Eclipse→New Project→wordcount→location /home/training/workspace
3. Right Click on the wordcount project→properties→java build path→Libraries→Add External Jars→Select all jars from /usr/lib/hadoop and /usr/lib/hadoop-0.20-mapreduce→Ok
4. Make sure that there are no more compilation errors

Step 3: Create a jar file

1. Right click the project→Export→Java→Jar→Select the location as /home/training→Make sure wordcount is checked→Finish

Step 4 Run the jar file

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