

LAB 3

This lab will get you started with MapReduce programming on UTD cluster.

1. You can download a working Maven project from here:

<http://www.utdallas.edu/~axn112530/cs6350/lab3/WordCount.zip>

2. Unzip it and import into your workspace as usual.

3. Build the maven project. Create the jar file.

4. Upload jar to your CS6360 directory, (not HDFS).

5. Your files from assignment 1 should be on HDFS already. So you should be able to run the following command:

```
hadoop jar WordCount.jar WordCount.WordCount.WordCount <Path to directories that you downloaded the files in> <Path to output directory>
```

****NOTE: OUTPUT DIRECTORY SHOULDN'T EXIST ON HDFS****

For example, I run the following:

```
hadoop jar WordCount.jar WordCount.WordCount.WordCount /user/axn112530/mrinput  
/user/axn112530/mroutput
```

NOTE: /user/axn112530/mroutput should NOT already exist.