

Setting up exercise

- Download Hadoop 0.20.2
- quit eclipse
- copy `hadoop-0.20.2/contrib/eclipse-plugin/hadoop-0.20.2-eclipse-plugin.jar` to `eclipse/plugins`
- restart eclipse
- checkout `exercises/HadoopExercise` from cvs

Exercise

- Download <http://www.cs.umd.edu/class/spring2010/cmssc433/shakespeare.tgz> or <http://www.cs.umd.edu/class/spring2010/cmssc433/shakespeare.zip>
- unpack it in the HadoopExercise directory
 - creates directory shakespeare

Exercise, continued

- Run WordCount with and without combiner
 - measure reduce input and output records
- Implement LineIndexer
 - measure reduce input and output records
 - try with and without combiner