How to Install Hadoop? (On Mac OS, Linux or Cygwin on Windows)

1) Download hadoop 0.20.0 from
   http://hadoop.apache.org/mapreduce/releases.html
2) Untar the hadoop file:
   ```
   tar xvfz hadoop-0.20.2.tar.gz
   ```
3) Set the path to java compiler by editing JAVA_HOME parameter in
   hadoop/conf/hadoop-env.sh:
   - Mac OS users can use
     /System/Library/Frameworks/JavaVM.framework/Versions/1.6.0/Home
   - Linux users can run "which java" command to obtain the path. Note that
     the JAVA_HOME variable shouldn't contain the bin/java at the end of
     path.
4) Create an RSA key to be used by hadoop when ssh'ing to localhost:
   ```
   ssh-keygen -t rsa -P ""
   cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
   ```
5) Do the following changes to the configuration files under hadoop/conf

   - **core-site.xml:**
   ```
   <configuration>
        <property>
             <name>hadoop.tmp.dir</name>
             <value>TEMPORARY-DIR-FOR-HADOOP-
   DATASTORE</value>
        </property>
        <property>
             <name>fs.default.name</name>
             <value>hdfs://localhost:54310</value>
        </property>
   </configuration>
   ```

   - **mapred-site.xml:**
   ```
   <configuration>
   <property>
             <name>mapred.job.tracker</name>
             <value>localhost:54311</value>
   </property>
   </configuration>
   ```

   - **hdfs-site.xml:**
   ```
   <configuration>
   <property>
             <name>dfs.replication</name>
             <value>1</value>
   </property>
   </configuration>
   ```

6) Format the hadoop file system. From hadoop directory run the following:
```
./bin/hadoop namenode -format
```
7) Run hadoop by running the following script:
```
./bin/start-all.sh
```
8) Now you can copy some data from your machine's file system into hdfs and do 'ls' command on hdfs:
```
./bin/hadoop dfs -put local_machine_path hdfs_path
./bin/hadoop dfs -ls
```
9) At this point you are ready to run a map reduce job on hadoop. As an example, let's run WordCount.jar to count the number of times each word appears in a text file. Put a sample text file on hdfs under 'input' directory. Download the jar file from:
http://www.stanford.edu/class/cs246/cs246-11-mmds/hw_files/WordCount.jar

and run the WordCount map-reduce job:

```
./bin/hadoop dfs -mkdir input
./bin/hadoop dfs -put local_machine_path/sample.txt input/sample.txt
./bin/hadoop jar ~/path_to_jar_file/WordCount.jar WordCount input
output
```

The result will be saved on 'output' directory on hdfs.

References:

http://arifn.web.id/blog/2010/07/29/running-hadoop-single-cluster.html
http://arifn.web.id/blog/2010/01/23/hadoop-in-netbeans.html
http://www.infosci.cornell.edu/hadoop/mac.html
http://wiki.apache.org/hadoop/GettingStartedWithHadoop