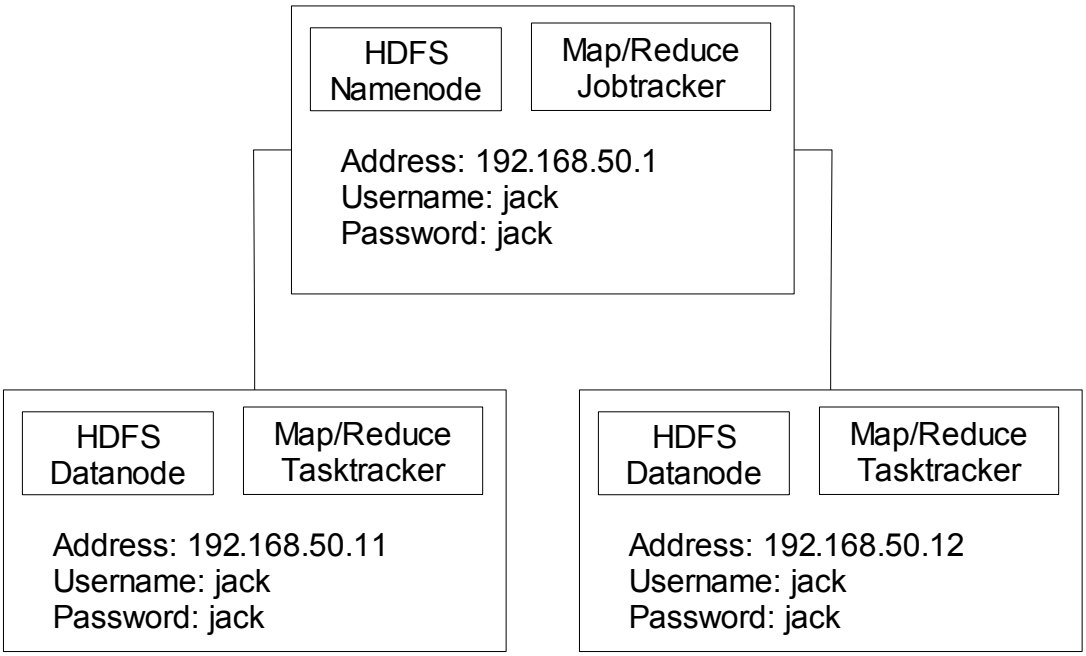
OpenSolaris / Hadoop Virtual Testbed Live CD

http://opensolaris.org/os/project/livehadoop/

Installation and usage instructions

1. Begin by placing the OpenSolaris / Hadoop Virtual Testbed Live CD into the CD/DVD drive of your computer. Turn on your computer, or if it is already on, restart it. During boot you may need to select the CD/DVD drive as the place to boot from. The way you do this varies from computer to computer. For Apple desktops and laptops, hold down the 'c' key during startup.
2. After a few minutes of booting, during which time you may see various status messages from the OpenSolaris operating system, you will be asked for the language to use during this session. Select the language and press return, or simply press return to accept the default of US-English.
3. After a minute or two you will be brought into a graphical Solaris environment. Read through the OpenSolaris Developer Preview License, then select 'Close' to dismiss the license window.
4. To begin using Hadoop, obtain a system console by selecting the 'Applications' menu from the top of the screen, then select 'System Tools', then select 'Terminal'.
5. You are now logged into the "Global" OpenSolaris zone that will run the Hadoop HDFS namenode and Map/Reduce jobtracker. You can find the Hadoop installation directory by typing:
    1. jack@opensolaris:~$ cd /opt/hadoop/hadoop-0.17.1
6. The environment included on the CD includes the global zone, as well as two "worker" zones that will run the actual Map/Reduce jobs and store datablocks for the HDFS filesystem. These zones are organized as follows:



7. For a simple demonstration of Map/Reduce, we are going to run the 'wordcount' demo that is included with the Hadoop distribution. We will use the contents of the /opt/hadoop/hadoop-0.17.1/conf directory as our input files. Once the demo is finished, a list of words and their count will be stored in the HDFS filesystem in the file /myoutput/part-00000

   To begin, type the following commands as shown here:

   1. jack@opensolaris:/opt/hadoop/hadoop-0.17.1$ bin/hadoop namenode -format
      - *This formats the HDFS file system on the namenode*
      - *From here on, we will omit the jack@opensolaris part of the prompt*
   2. $ bin/start-dfs.sh
   3. $ bin/start-mapred.sh
      - *These commands bring up the HDFS and Map/Reduce layers*
   4. $ zoneadm list
      - This shows you the zones are running in OpenSolaris. 'node1' and 'node2' are worker zones. You can log into them by typing either "ssh 192.168.50.11" or "ssh 192.168.50.12"
   5. $ bin/hadoop fs -mkdir /myinput
      - *Creates an input directory that will hold our sample input*
   6. $ bin/hadoop fs -ls /myoutput
      - *This command will cause an error, indicating that "/myoutput" does not exist. That is ok, because the wordcount program will create this directory for you automatically in step 8.*
   7. $ bin/hadoop fs -copyFromLocal conf/*.xml /myinput
      - *Populates our input directory with sample input, in this case a few configuration files*
   8. $ bin/hadoop jar hadoop-0.17.1-examples.jar wordcount /myinput /myoutput
      - *This processes the wordcount demo and stores the result in HDFS in the*
      - *file '/myoutput/part-00000'*
   9. $ bin/hadoop fs -cat /myoutput/part-00000
      - *This command lets you examine the result of the Map/Reduce job*
   10. $ bin/stop-mapred.sh
   11. $ bin/stop-dfs.sh
      - *These commands shut down Hadoop by stopping the HDFS and Map/Reduce layers*