Steps in Installing HBase

Step 1- Take 2 laptops. Install connectify software in both the laptops. Connectify creates a private wireless lan between 2 or more laptops. There is a ip assigned to each of the machines by this connectify. We will use these ips to setup hadoop and eventually hbase.

Alternatively, ubuntu has the option for creating a Wi-Fi network by default which can be accessed by clicking on the network icon on the top of the screen. Once Wi-Fi is setup and the laptops are connected by this network, use the ifconfig command in the terminal to get the IP addresses which are to be used for these installations.

Step 2 - Install Hadoop in each of these machines separately using the documentation given in the following link -
http://www.michael-noll.com/tutorials/running-hadoop-on-ubuntu-linux-single-node-cluster/

These steps are to be followed before formatting the hdfs file system via the namenode:
1. Make a directory at any location on your system (eg. Say a directory dfs is made in home directory)
2. Create two subdirectories in this directory (eg. named data and name) in this directory.
3. Add the following lines to /usr/local/hadoop/conf/hdfs-site.xml file between the <configuration></configuration>:

```
<property>
    <name>dfs.name.dir</name>
    <value>/home/hduser/dfs/name/</value>
    <description> namenode directory. </description>
</property>

<property>
    <name>dfs.data.dir</name>
    <value>/home/hduser/dfs/data/</value>
    <description> namenode directory. </description>
</property>
```

Some places where problems might arise:

1. Update $HOME/.bashrc : while adding the lines mentioned in this step described by the above documentation, to find the JAVA_HOME path on your system, use the command
 readlink -f $(which java)

Make the appropriate changes in the export JAVA_HOME line of ~/.bashrc file. The same path is to be used to set JAVA_HOME path in /usr/local/hadoop/conf/hadoop-env.sh file.

2. The jps command may give an output as follows:

```
The program 'jps' can be found in the following packages:
 * openjdk-7-jdk
 * openjdk-6-jdk
Try: sudo apt-get install <selected package>
```

For this use the following command:

```
alias jps = '/usr/local/java/jdk1.8.0_05/bin/jps'
```

The above path is an example, it might differ from system to system based on java installation settings. The alias is to be set to 'path/bin/jps' where path is the installed jdk directory.

After installing hadoop on each of the node, now follow the instructions in the following documentation to create **a multi node hadoop cluster.**
http://www.michael-noll.com/tutorials/running-hadoop-on-ubuntu-linux-multi-node-cluster/

An extra step to be followed while updating /etc/hosts file:
Add the following line to the /etc/hosts file in the master node and slave nodes:

```
192.168.0.1      namenode
```

where 192.168.0.1 is the IP of the machine being used as the master

Please use **hadoop version 1.0.3**. The setup can be downloaded from the link -
http://172.24.2.57/hadoop-1.0.3.tar.gz

This is just a temporary link and shall be deactivated very soon.

Step 3 - After installing hadoop cluster, install hbase as per the documentation given in the following link -
http://jayatiatblogs.blogspot.in/2013/01/hbase-installation-fully-distributed.html

Please use **Hbase version - 0.94.x**. The setup can be downloaded from the link -
http://172.24.2.57/hbase-0.94.18.tar.gz

Please use the versions mentioned above only. Not all versions of hadoop are compatible with all versions of hbase. Also, the documentation by Micheal Noll for hadoop is only for the version 1.0.3. Higher versions have different steps for installation.

Courtesy: Jagat Sesh, Nitin