**Creating two node Hadoop cluster (Master and Slave)**

Host OS: Windows 7

BaseBox: Ubuntu (Lucid32)

Technology: Oracle VirtualBox and Vagrant

1. Prerequisites
   1. Install Vagrant Tool
   2. Install VirtualBox (<http://www.virtualbox.org>) latest version
   3. Put the Virutalbox installation path to the environment variable (PATH)
   4. Download Virtualbox base box from the [www.virutalboxes.es](http://www.virutalboxes.es)
2. Pick a folder for your work. Assuming D:\vagrant\hadoop for the purpose of this documentation.
   1. Work in this folder from now on.
3. Create VirtualMachines from BaseBox (Lucid32)
   1. Create “master” base box
      1. Vagrant box add “master” “<Path to the Lucid32.box>”
      2. Vagrant init
   2. Create “slave” base box
      1. Vagrant box add “slave” “<Path to the Lucid32.box>”
      2. Vagrant init
4. Folder will now have a vagrantfile. Replace the vagrant file with our vagrant file.
   1. Our vagrant file will start the master as “Hadoop Master” and slave as “Hadoop Slave”
   2. Hadoop master will have the IP: 10.10.10.50 and slave will have 10.10.10.51
5. Do “vagrant up”
   1. Both master and slave machine should be up and running and should be visible in VirtualBox UI as “Hadoop Master” and “Hadoop Slave”
6. Logging into the machines
   1. On Windows you can use Putty-Gen to create a private key (.ppk) file from the Private key which is in .vagrant.d folder
   2. You can access the machines with the IP we have given.
   3. You will get the shell when you login.
7. Do the following on both Master and Slave machines to install JDK and Hadoop
   1. Create a folder ~/bin and copy the files init.sh, jdk.sh and Hadoop.sh into this folder.
   2. Run init.sh
   3. Run jdk.sh
   4. Run Hadoop.sh
   5. To the file ~/hadoop/conf/Hadoop-env.sh file add the line
      1. “export JAVA\_HOME=/usr/lib/jvm/java-7-oracle”
8. Now you are ready to configure Hadoop on both machines.
9. Changes on Master
   1. Edit file ~hadoop/conf/masters
      1. Remove localhost
      2. Add line “master”
   2. Edit file ~hadoop/conf/slaves
      1. Remove localhost
      2. Add line “slave”
   3. Edit file .ssh/id\_rsa
      1. Add the data in private key that will be there in .vagrant/d folder
      2. This will allow auto login from master to slave.
   4. Edit /etc/hostname file and change “localhost” to “master”
   5. Edit /etc/hosts file
      1. Remove localhost lines.
      2. Add line “10.10.10.50 master”
      3. Add line “10.10.10.51 slave”
10. Changes on Slave
    1. Edit file ~hadoop/conf/masters
       1. Remove localhost
    2. Edit file ~hadoop/conf/slaves
       1. Remove localhost
       2. Add line “slave”
    3. Edit /etc/hostname file and change “localhost” to “slave”
    4. Edit /etc/hosts file
       1. Remove localhost lines.
       2. Add line “10.10.10.51 slave”
11. Do the following commands
    1. Vagrant halt
    2. Vagrant up
12. Login to Master machine
    1. Create folder ~/HDFS
    2. Run the command “hadoop namenode –format”
       1. This will create HDFS under ~/HDFS folder
    3. Run command “start-dfs.sh”
       1. This will start daemons
          1. NameNode and SecondaryNameNode on master
          2. DataNode on slave
    4. Run command “start-mapred.sh”
       1. This will start daemons
          1. JobTracker on master
          2. TaskTracker on slave
13. Running command “jps” on master should give the three master daemons
14. Running command “jps on slave should give the two slave daemons.
15. On master
    1. Run command “hadoop fs –put <filename> /” to put a file in HDFS
    2. Run command “hadoop fs –ls /” to check if you can see the uploaded file in HDFS.
16. Two node Hadoop cluster is configured now on two VMs
17. Using “vagrant package” command to create your own BaseBoxes “hmaster” and “hslave” to create another Hadoop cluster.
    1. Delete downloaded tar/gz files JDK and Hadoop from both master and slave.
    2. Run command “vagrant package master –vagrantfile Vagrantfile –output hmaster.box
    3. Run command “vagrant package slave –vagrantfile Vagrantfile –output hslave.box
    4. Now you can use these BaseBoxes and create two other VMs that will have Hadoop installed and configured on it.