Batch-L10

Roll number-33235

Multi Node Cluster in Hadoop.

## Prerequisites

* Cent OS 6.5
* Hadoop-2.7.3
* JAVA 8
* SSH

## Setup of Multi Node Cluster in Hadoop

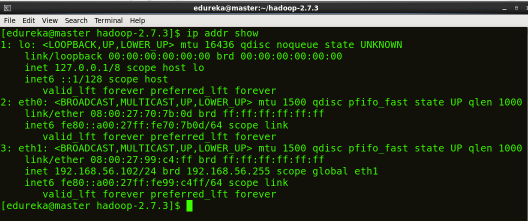
We have two machines (master and slave) with IP:

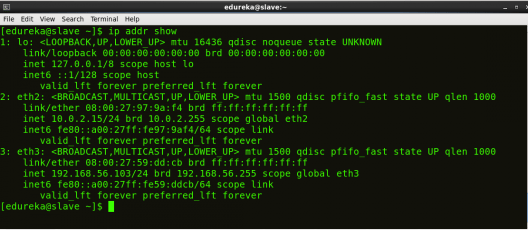
Master IP: 192.168.56.102

Slave IP: 192.168.56.103

STEP 1: Check the IP address of all machines.

Command: ip addr show (you can use the ifconfig command as well)





STEP 2: Disable the firewall restrictions.

Command: service iptables stop

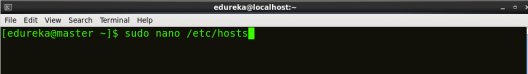
Command: sudo chkconfig iptables off

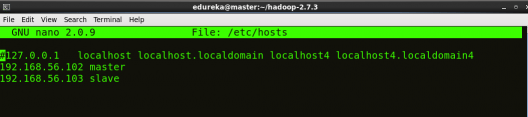


STEP 3: Open hosts file to add master and data node with their respective IP addresses.

Command: sudo nano /etc/hosts

Same properties will be displayed in the master and slave hosts files.





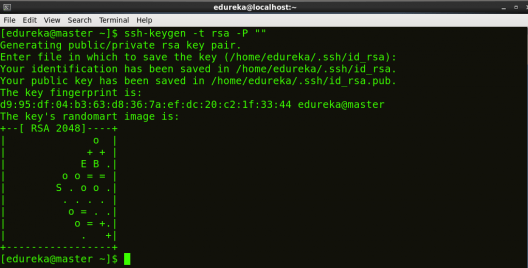
STEP 4: Restart the sshd service.

Command: service sshd restart



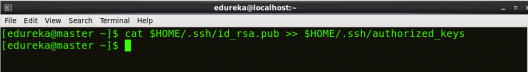
STEP 5: Create the SSH Key in the master node. (Press enter button when it asks you to enter a filename to save the key).

Command: ssh-keygen -t rsa -P “”



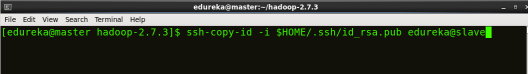
STEP 6: Copy the generated ssh key to master node’s authorized keys.

Command: cat $HOME/.ssh/id\_rsa.pub >> $HOME/.ssh/authorized\_keys



STEP 7: Copy the master node’s ssh key to slave’s authorized keys.

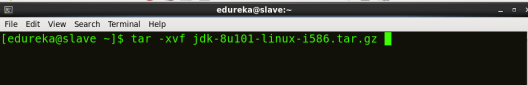
Command: ssh-copy-id -i $HOME/.ssh/id\_rsa.pub edureka@slave



STEP 8: [Click here](https://goo.gl/B2BAk2) to download the Java 8 Package. Save this file in your home directory.

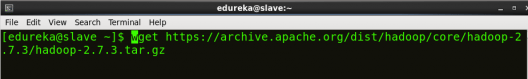
STEP 9: Extract the Java Tar File on all nodes.

Command: tar -xvf jdk-8u101-linux-i586.tar.gz



STEP 10: Download the Hadoop 2.7.3 Package on all nodes.

Command: wget https://archive.apache.org/dist/hadoop/core/hadoop-2.7.3/hadoop-2.7.3.tar.gz



STEP 11: Extract the Hadoop tar File on all nodes.

Command: tar -xvf hadoop-2.7.3.tar.gz

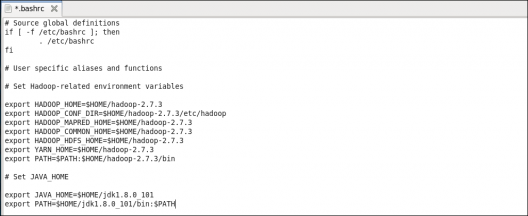


STEP 12: Add the Hadoop and Java paths in the bash file (.bashrc) on all nodes.

Open .bashrc file. Now, add Hadoop and Java Path as shown below:

Command: sudo gedit .bashrc

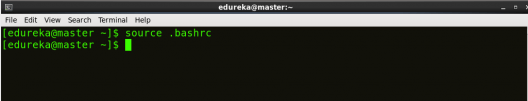




Then, save the bash file and close it.

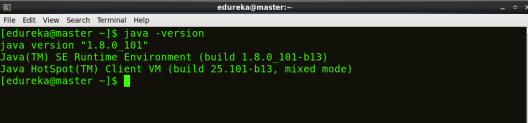
For applying all these changes to the current Terminal, execute the source command.

Command: source .bashrc

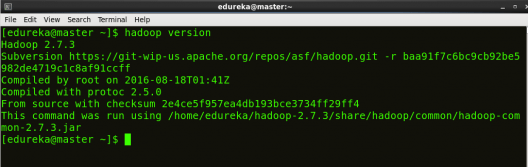


To make sure that Java and Hadoop have been properly installed on your system and can be accessed through the Terminal, execute the java -version and hadoop version commands.

Command: java-version



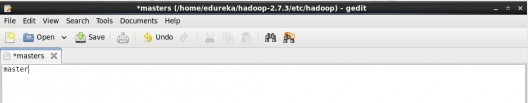
Command: hadoop version



Now edit the configuration files in hadoop-2.7.3/etc/hadoop directory.

STEP 13: Create masters file and edit as follows in both master and slave machines as below:

Command: sudo gedit masters



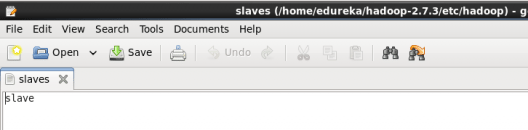
STEP 14: Edit slaves file in master machine as follows:

Command: sudo gedit /home/edureka/hadoop-2.7.3/etc/hadoop/slaves



STEP 15: Edit slaves file in slave machine as follows:

Command: sudo gedit /home/edureka/hadoop-2.7.3/etc/hadoop/slaves



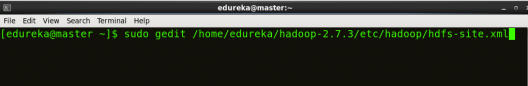
STEP 16: Edit core-site.xml on both master and slave machines as follows:

Command: sudo gedit /home/edureka/hadoop-2.7.3/etc/hadoop/core-site.xml



|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8 | &lt;?xml version="1.0" encoding="UTF-8"?&gt;  &lt;?xml-stylesheet type="text/xsl" href="configuration.xsl"?&gt;  &lt;configuration&gt;  &lt;property&gt;  &lt;name&gt;fs.default.name&lt;/name&gt;  &lt;value&gt;hdfs://master:9000&lt;/value&gt;  &lt;/property&gt;  &lt;/configuration&gt; |

STEP 17: Edit hdfs-site.xml on master as follows:  
Command: sudo gedit /home/edureka/hadoop-2.7.3/etc/hadoop/hdfs-site.xml



|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18  19  20 | &lt;?xml version="1.0" encoding="UTF-8"?&gt;  &lt;?xml-stylesheet type="text/xsl" href="configuration.xsl"?&gt;  &lt;configuration&gt;  &lt;property&gt;  &lt;name&gt;dfs.replication&lt;/name&gt;  &lt;value&gt;2&lt;/value&gt;  &lt;/property&gt;  &lt;property&gt;  &lt;name&gt;dfs.permissions&lt;/name&gt;  &lt;value&gt;false&lt;/value&gt;  &lt;/property&gt;  &lt;property&gt;  &lt;name&gt;dfs.namenode.name.dir&lt;/name&gt;  &lt;value&gt;/home/edureka/hadoop-2.7.3/namenode&lt;/value&gt;  &lt;/property&gt;  &lt;property&gt;  &lt;name&gt;dfs.datanode.data.dir&lt;/name&gt;  &lt;value&gt;/home/edureka/hadoop-2.7.3/datanode&lt;/value&gt;  &lt;/property&gt;  &lt;/configuration&gt; |

STEP 18: Edit hdfs-site.xml on slave machine as follows:

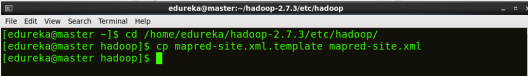
Command: sudo gedit /home/edureka/hadoop-2.7.3/etc/hadoop/hdfs-site.xml

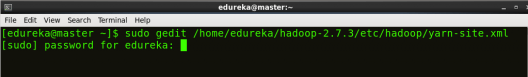
|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16 | &lt;?xml version="1.0" encoding="UTF-8"?&gt;  &lt;?xml-stylesheet type="text/xsl" href="configuration.xsl"?&gt;  &lt;configuration&gt;  &lt;property&gt;  &lt;name&gt;dfs.replication&lt;/name&gt;  &lt;value&gt;2&lt;/value&gt;  &lt;/property&gt;  &lt;property&gt;  &lt;name&gt;dfs.permissions&lt;/name&gt;  &lt;value&gt;false&lt;/value&gt;  &lt;/property&gt;  &lt;property&gt;  &lt;name&gt;dfs.datanode.data.dir&lt;/name&gt;  &lt;value&gt;/home/edureka/hadoop-2.7.3/datanode&lt;/value&gt;  &lt;/property&gt;  &lt;/configuration&gt; |

STEP 19: Copy mapred-site from the template in configuration folder and the edit mapred-site.xml on both master and slave machines as follows:

Command: cp mapred-site.xml.template mapred-site.xml

Command: sudo gedit /home/edureka/hadoop-2.7.3/etc/hadoop/mapred-site.xml

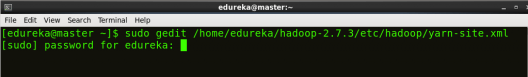




|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8 | &lt;?xml version="1.0" encoding="UTF-8"?&gt;  &lt;?xml-stylesheet type="text/xsl" href="configuration.xsl"?&gt;  &lt;configuration&gt;  &lt;property&gt;  &lt;name&gt;mapreduce.framework.name&lt;/name&gt;  &lt;value&gt;yarn&lt;/value&gt;  &lt;/property&gt;  &lt;/configuration&gt; |

STEP 20: Edit yarn-site.xml on both master and slave machines as follows:

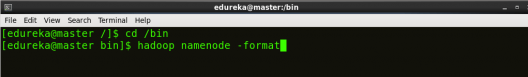
Command: sudo gedit /home/edureka/hadoop-2.7.3/etc/hadoop/yarn-site.xml



|  |  |
| --- | --- |
| 1  2  3  4  5  6  7  8  9  10  11  12 | &lt;?xml version="1.0" encoding="UTF-8"?&gt;  &lt;?xml-stylesheet type="text/xsl" href="configuration.xsl"?&gt;  &lt;configuration&gt;  &lt;property&gt;  &lt;name&gt;yarn.nodemanager.aux-services&lt;/name&gt;  &lt;value&gt;mapreduce\_shuffle&lt;/value&gt;  &lt;/property&gt;  &lt;property&gt;  &lt;name&gt;yarn.nodemanager.auxservices.mapreduce.shuffle.class&lt;/name&gt;  &lt;value&gt;org.apache.hadoop.mapred.ShuffleHandler&lt;/value&gt;  &lt;/property&gt;  &lt;/configuration&gt; |

STEP 21: Format the namenode(Only on master machine).

Command: hadoop namenode -format



STEP 22: Start all daemons(Only on master machine).

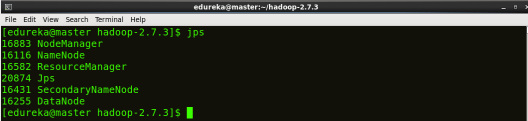
Command: ./sbin/start-all.sh



STEP 23: Check all the daemons running on both master and slave machines.

Command: jps

On master



On slave

