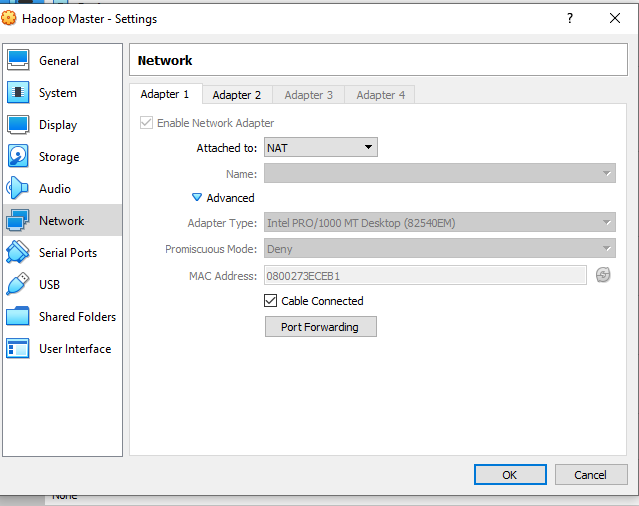
**Hadoop Multi node cluster setup**

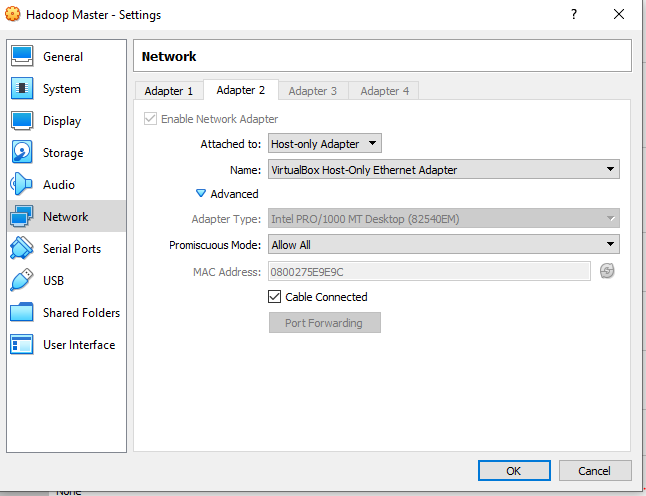
**Prerequisites:**

Ubuntu 18.04

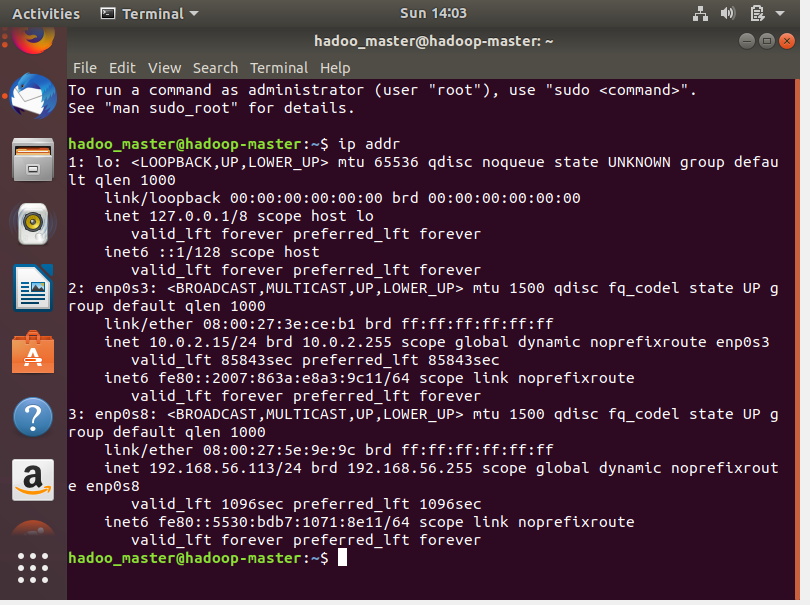
**Step 1 : Network Configuration**

1. Create a Virtual machine –“ Hadoop Master” on Oracle VM Virtual box. The network settings are as below:

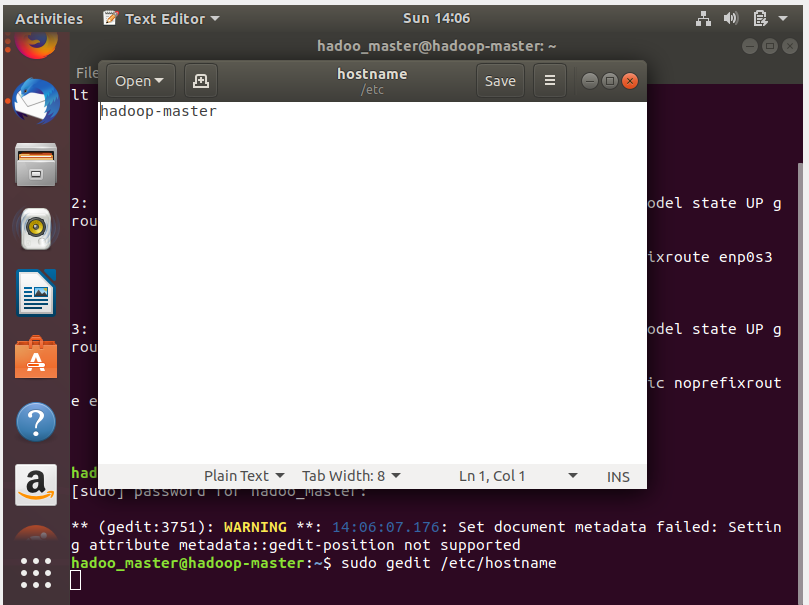




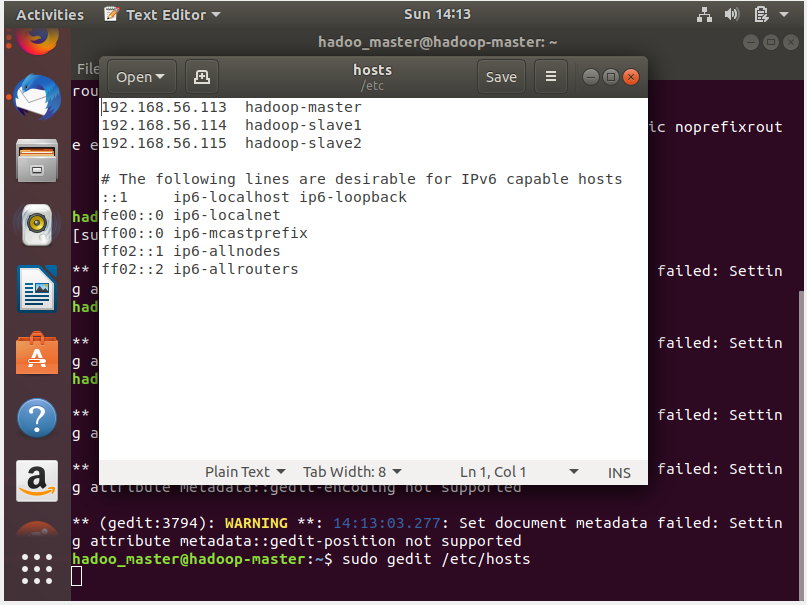
1. Get the ip address of Hadoop Master



1. /etc/hostname

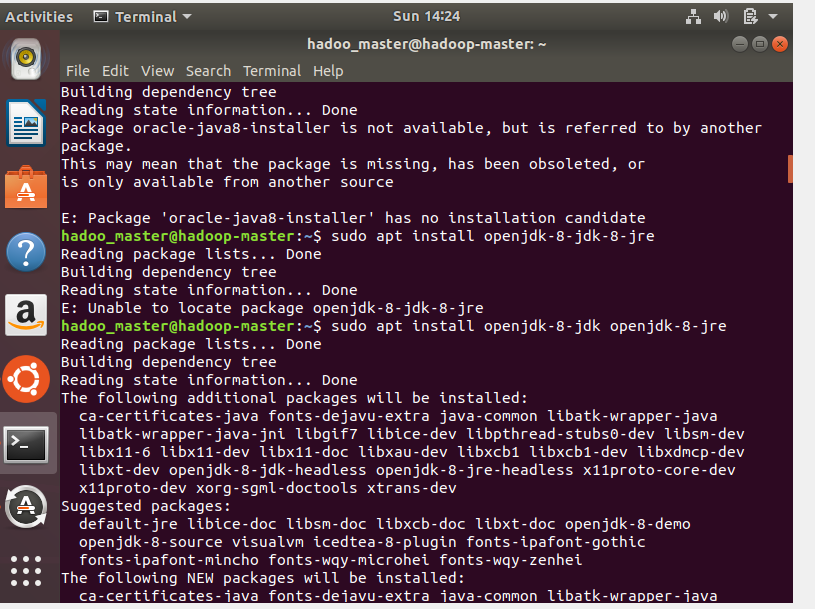
Edit 

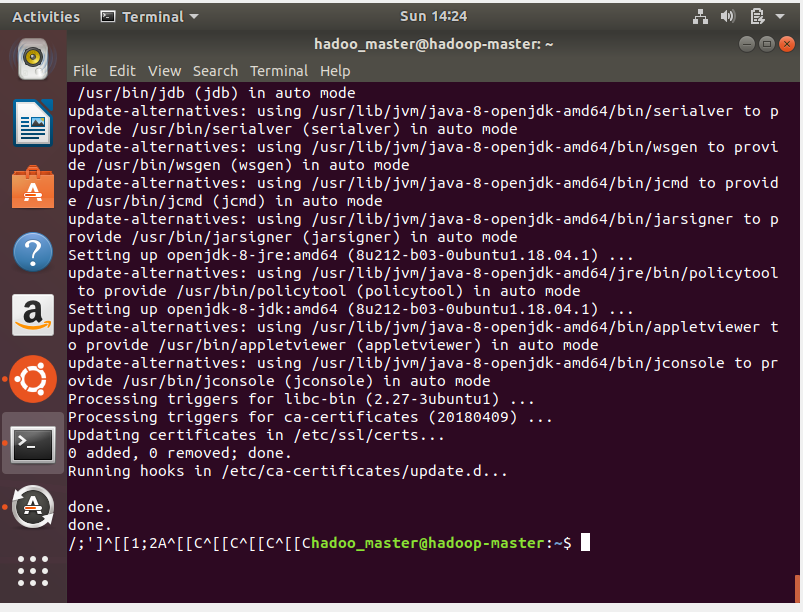
1. Edit /etc/hosts

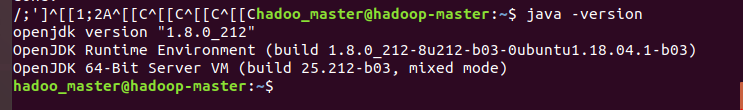


**Step 2 - Preparing the Hadoop Servers**

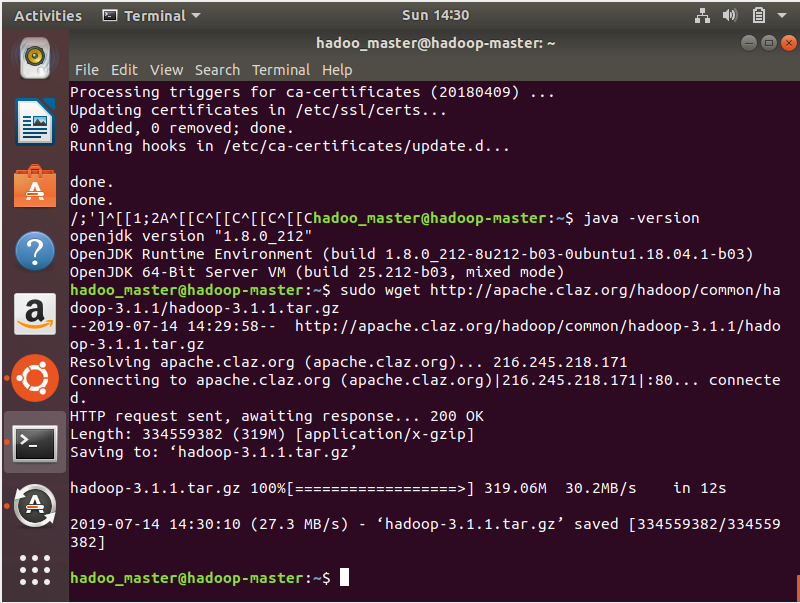
1. Java installation



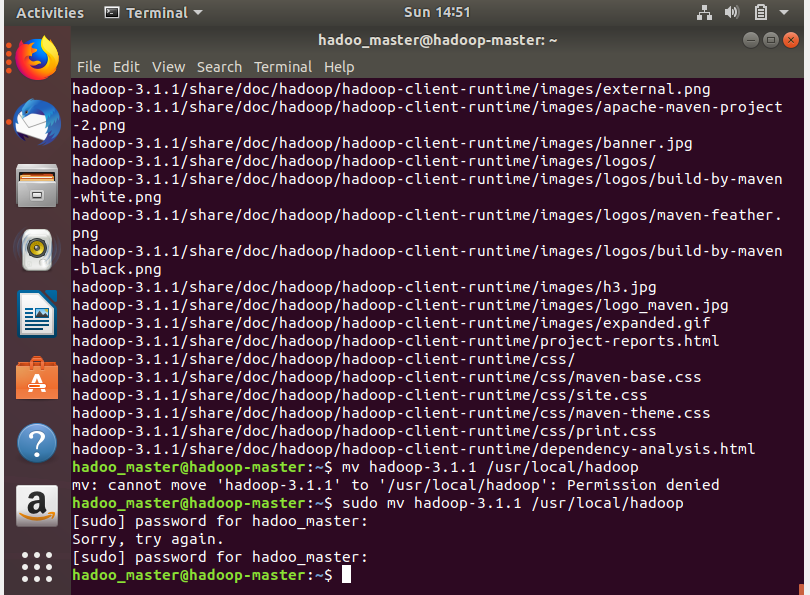




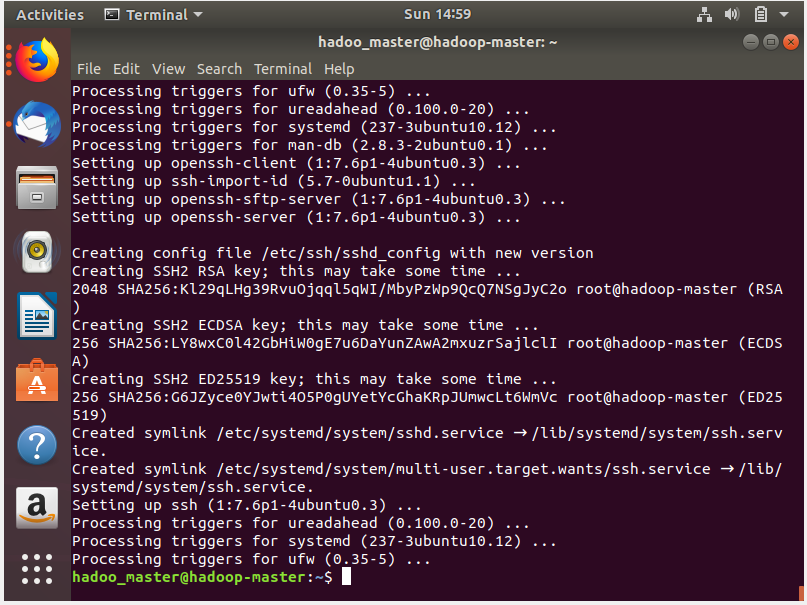
1. Download Hadoop



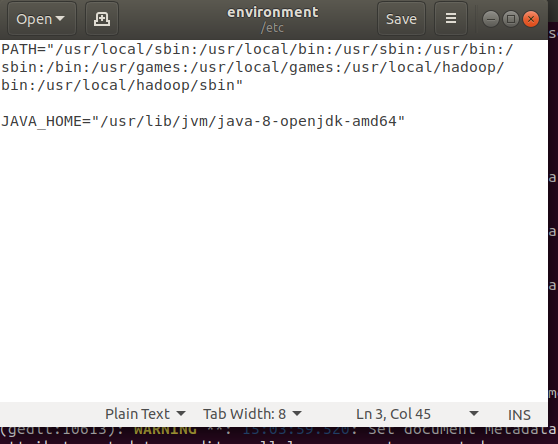
1. Untar the archive and move it to /usr/local.



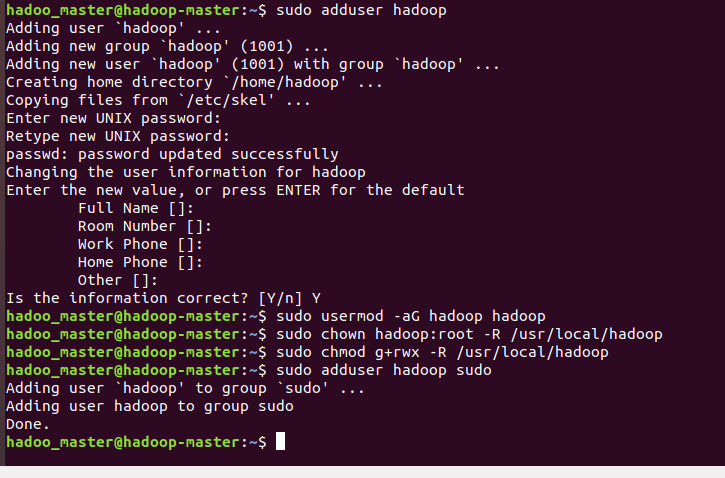
1. Install ssh



1. open /etc/environment and update the PATH line to include the Hadoop binary directories and also add a line for the JAVA\_HOME variable.

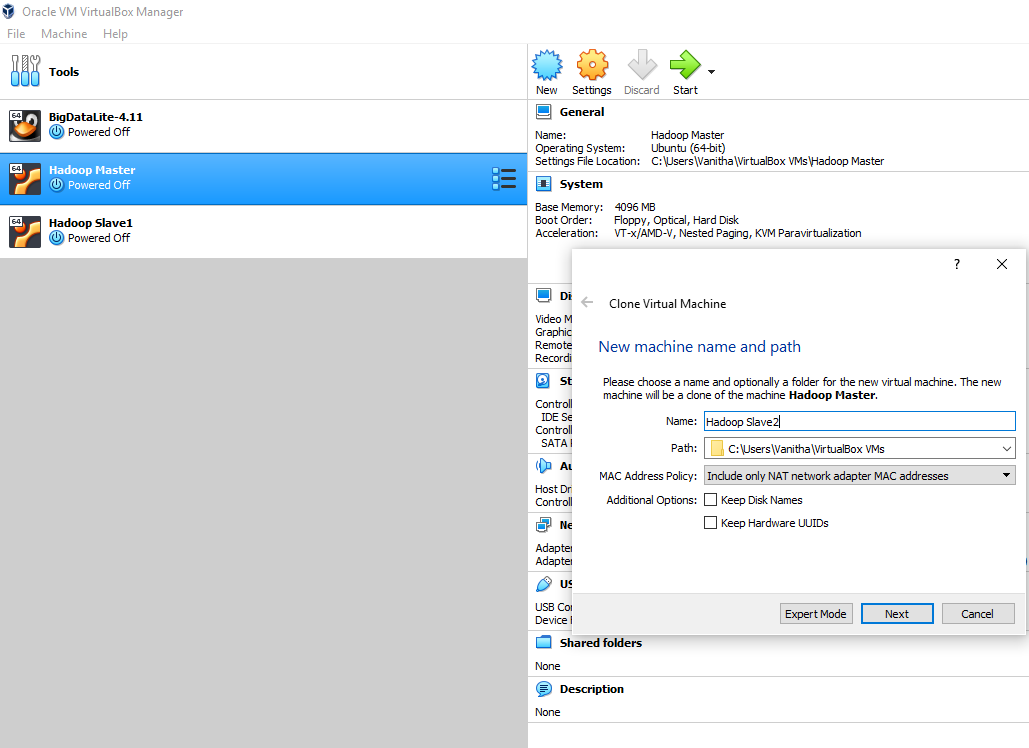


1. add a hadoop user and give him the necessary permissions

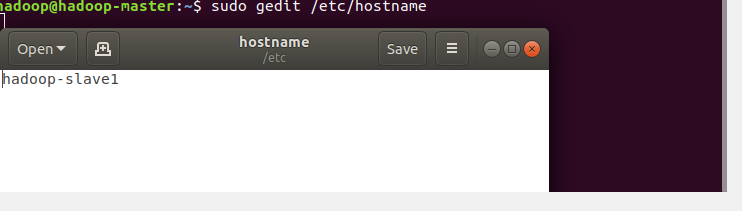


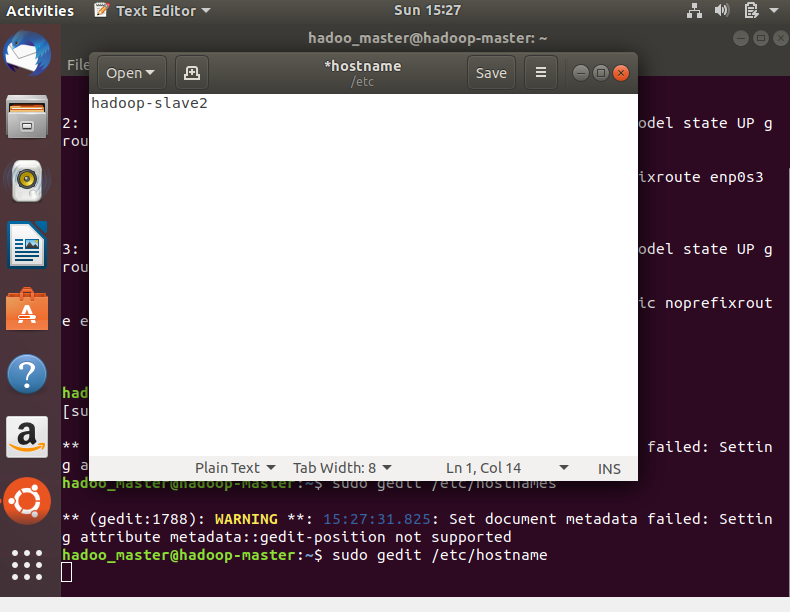
# Step 3 - Cloning the hadoop-master

1. clone the Hadoop Master to create the Hadoop Slave1 and the Hadoop Slave2



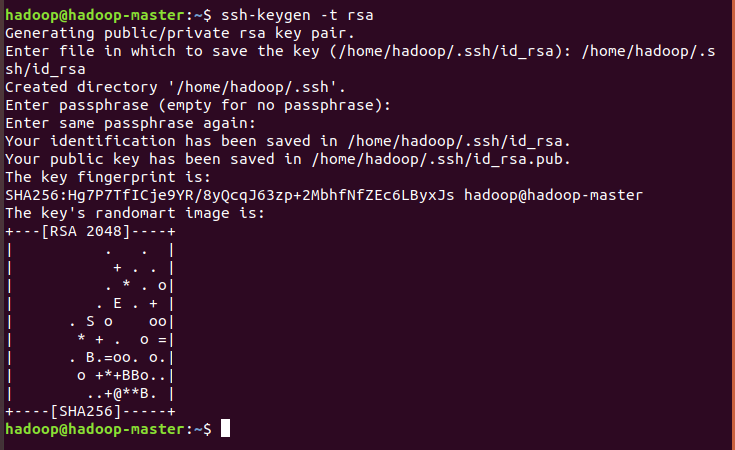
1. configure the hostname of the Slaves



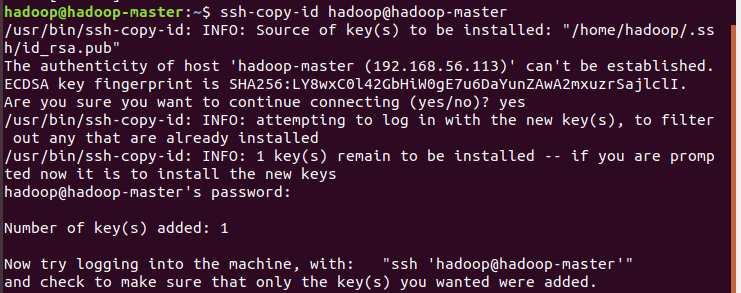


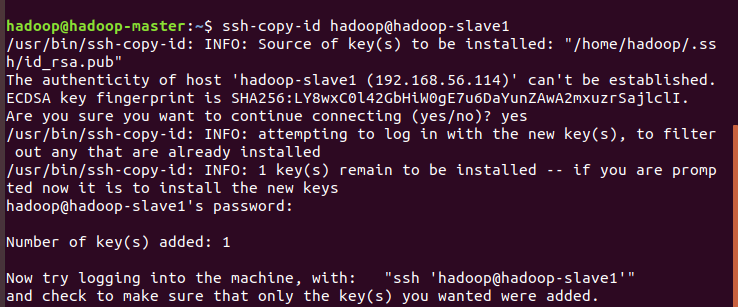
# Step 4 - SSH Keys

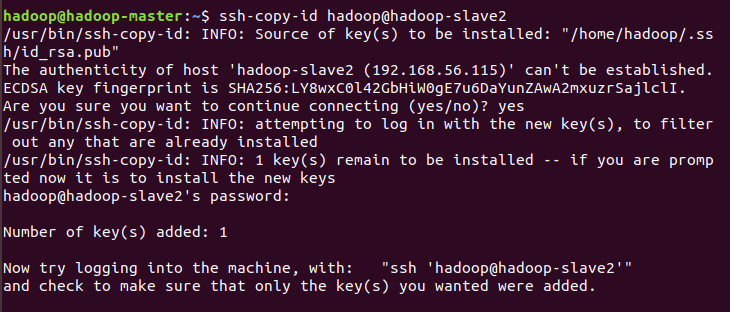
1. generate an SSH key.



1. Copy SSH keys to all users

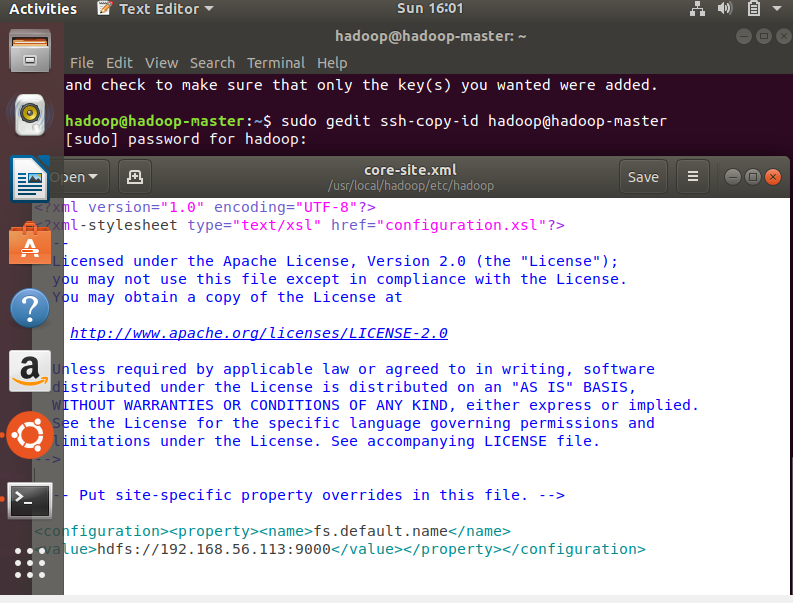




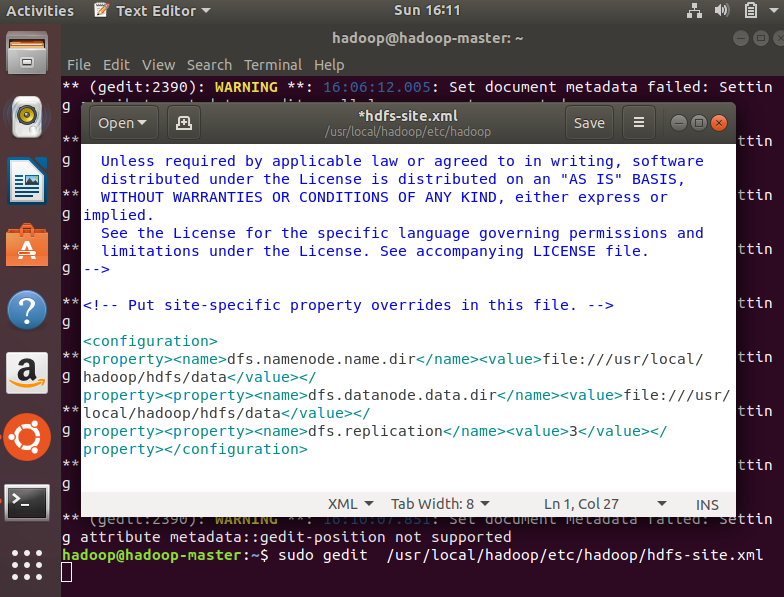


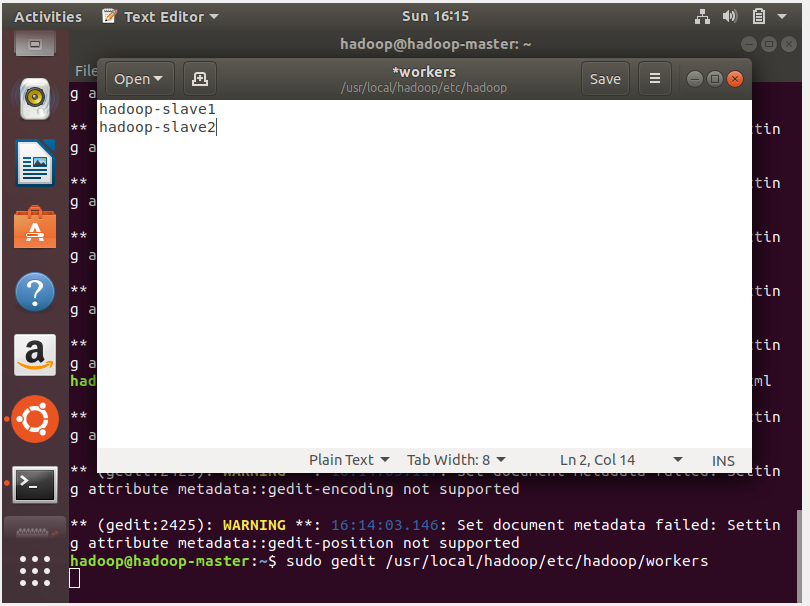
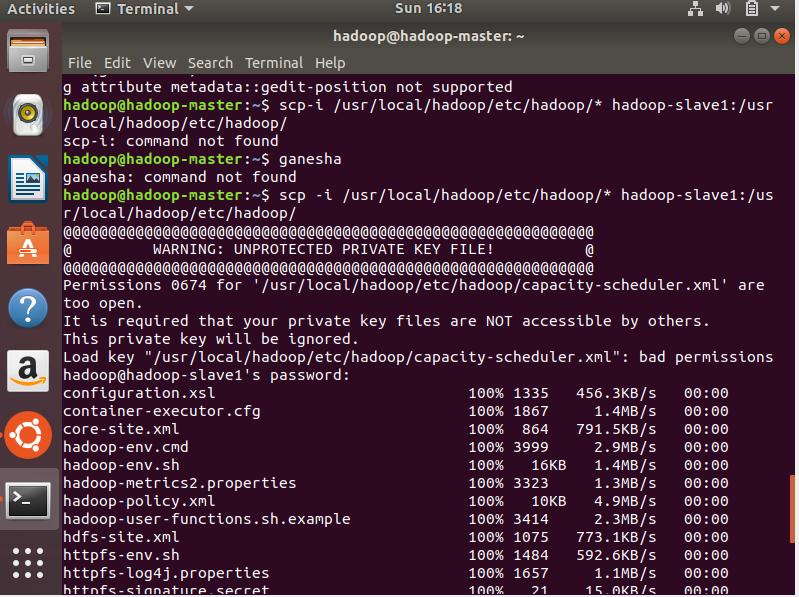
# Step 5 - Configuring the Hadoop Master

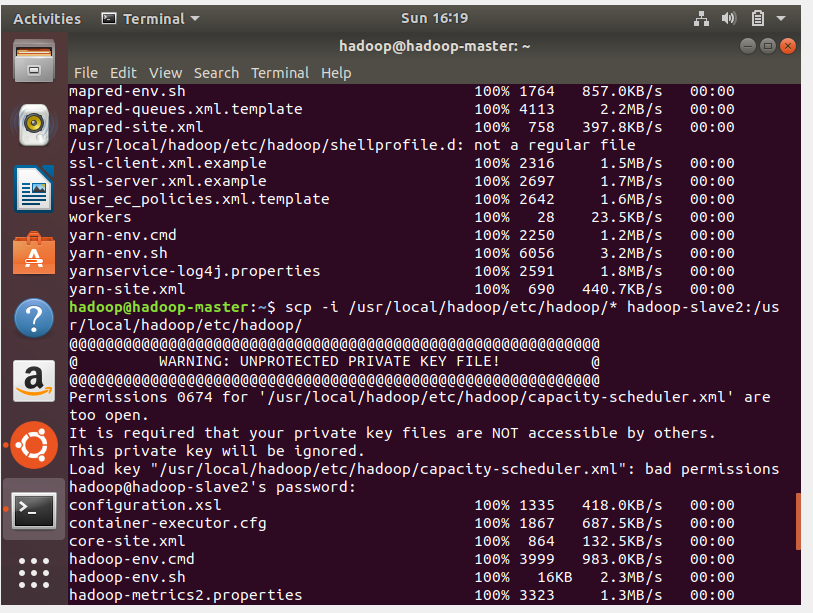
1. Edit core-site.xml file



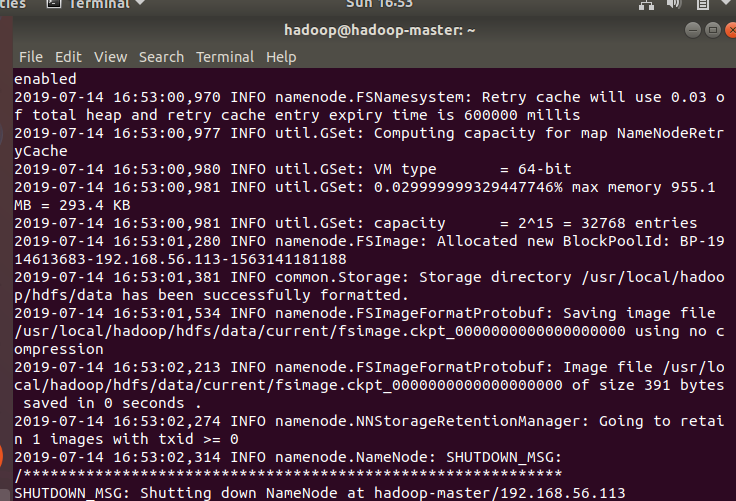
16. open the /usr/local/hadoop/etc/hadoop/hdfs-site.xml file and add the following:



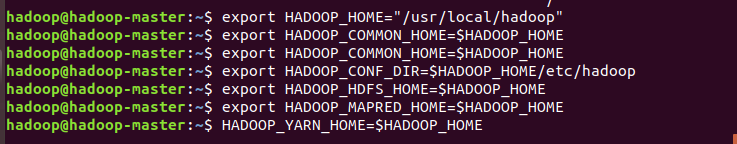
1. Open the /usr/local/hadoop/etc/hadoop/workers file and add the slaves hostname
2. Copy the configuration files from the Hadoop Master to the Slaves



1. Format HDFS Namenode

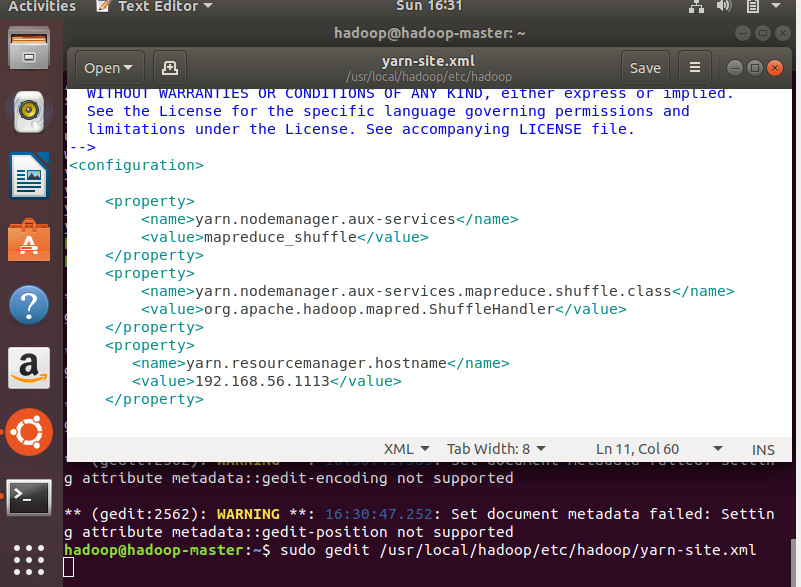


1. Execute the following commands

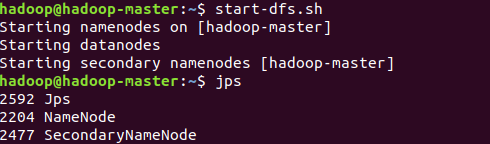


**Step 6 - Configuring Hadoop Slaves and Starting Hadoop**

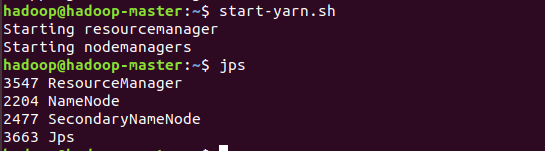
1. Edit **/usr/local/hadoop/etc/hadoop/yarn-site.xml** file at the Hadoop Slaves



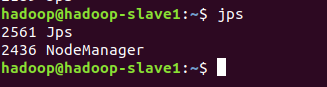
1. Start Hadoop on Hadoop Master and validate by running JPS command



1. Start Yarn on Hadoop Master and validate by running JPS command



1. Observe the output on slaves





**References**

1. <https://www.edureka.co/blog/apache-hadoop-hdfs-architecture/>
2. <https://medium.com/@ryannel/>[installing-multi-node-hadoop-on-ubuntu-7decd4c285f0](https://medium.com/@ryannel/installing-multi-node-hadoop-on-ubuntu-7decd4c285f0)
3. <https://dzone.com/articles/install-a-hadoop-cluster-on-ubuntu-18041>