**Session 5: MapReduce**

**Assignment 1: In this assignment you need to install single node Hadoop installation.**

# Installing Oracle Virtual Box

Download and install the Oracle Virtual Box from the below link.

**Link:** https://drive.google.com/file/d/0Bxr27gVaXO5sRXdxQVpEUmhCZ3c/view?usp=sharing

# Downloading CentOS

Download and install CentOS from the below link:

**Link:** https://drive.google.com/file/d/0Bxr27gVaXO5sRU0yVFVQM0FvLU0/view?usp=sharing

This is compressed file we need to unzip it.

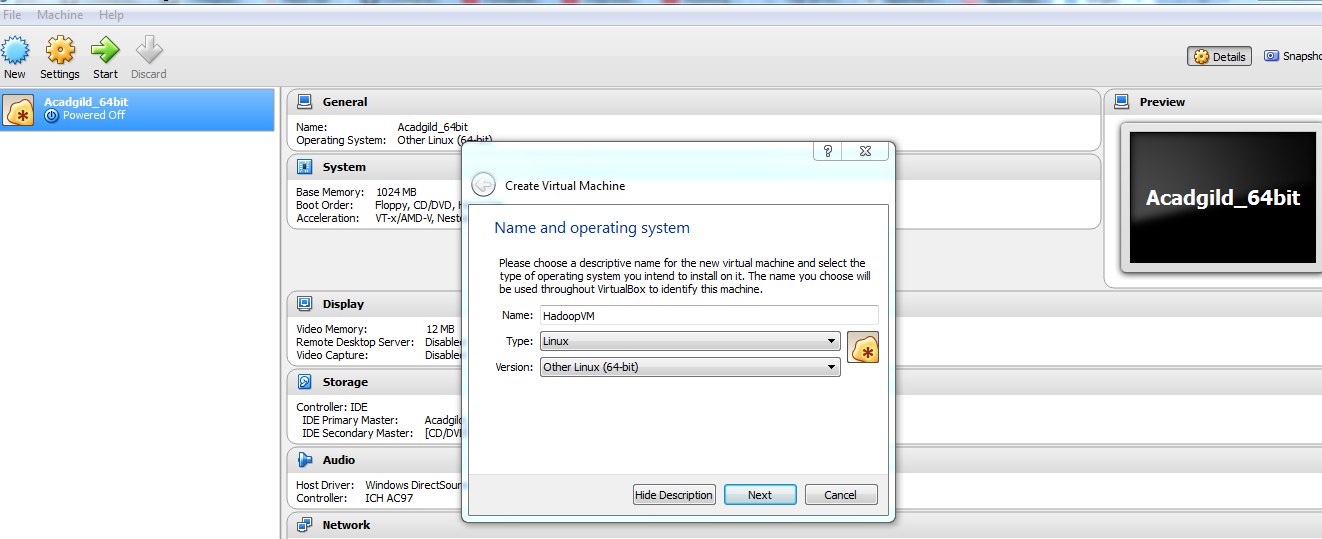
# Steps to Install CentOS:

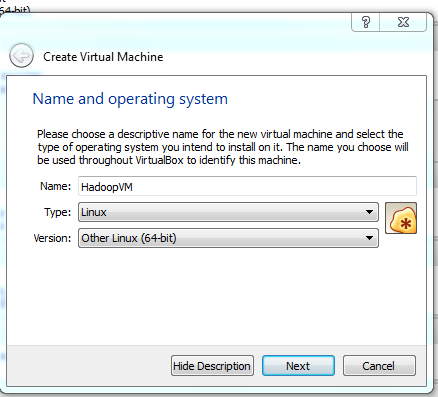
1. Click on New Option and then enter the below details as shown in the screenshot.Awindow pop up fill all the details

**Name:** Type in any name, to name your VM.(HadoopVM)

**Type:** Select theoption **Linux**from the drop down list.

**Version:** Select**Other Linux (64 bit)** from the drop down list.

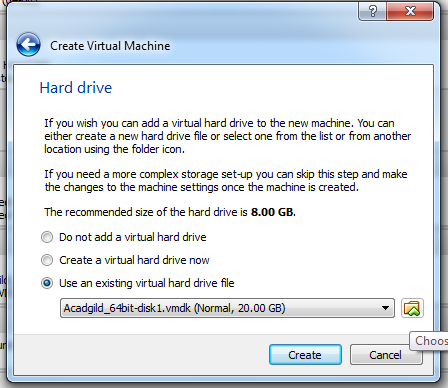
****

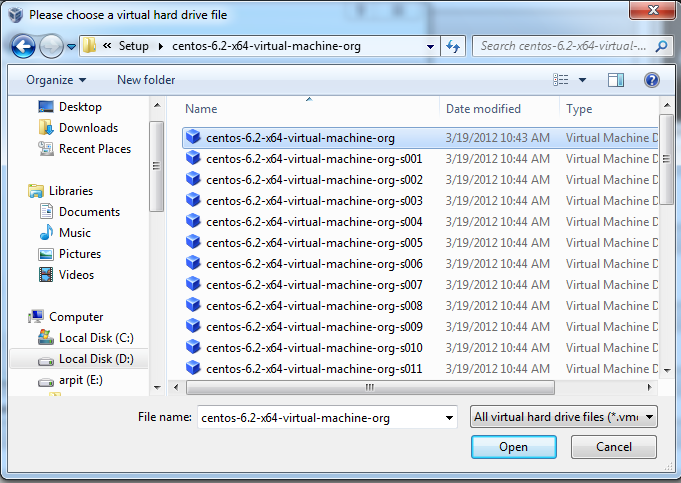
****

1. Click Next.

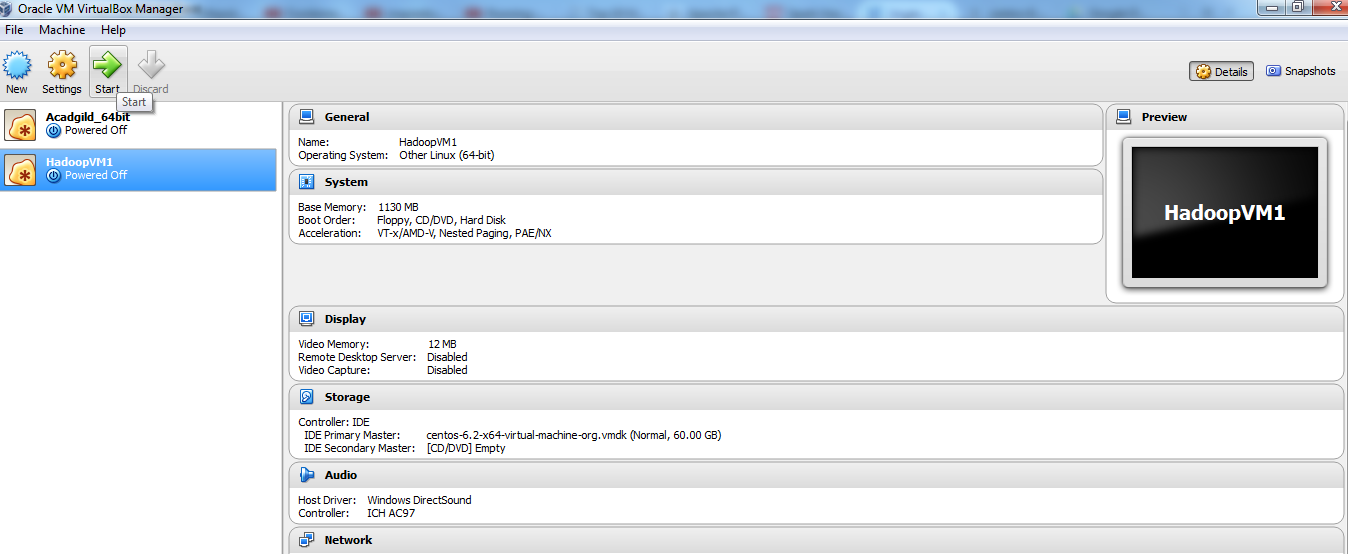
On clicking **Next**, a prompt appears to set up RAM size for the VM.

1. Increase the RAM up to 2048 MB if the system has 8 GB RAM and increase up to 1 GB if the system has 4 GB RAM.
2. Click on **Next** to get the option of selecting the Hard Disk Option; choose the third option i.e. using the existing Virtual hard drive file.
3. Click on the **folder icon** to browse to the location where the unzipped file of CentOS is kept.

****

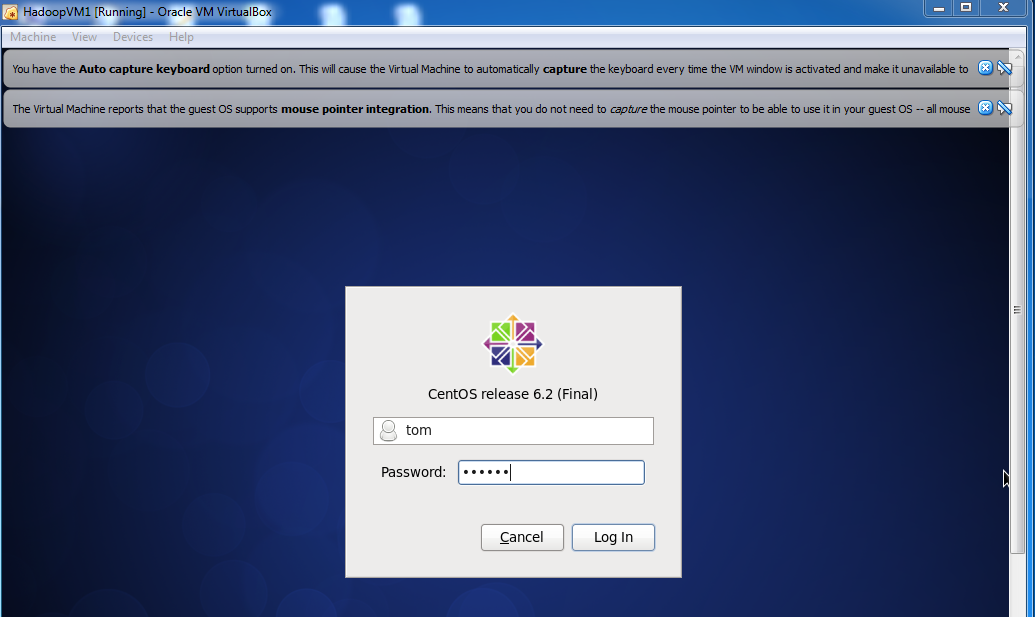
****

On starting the VM, a prompt appears to put the credentials.

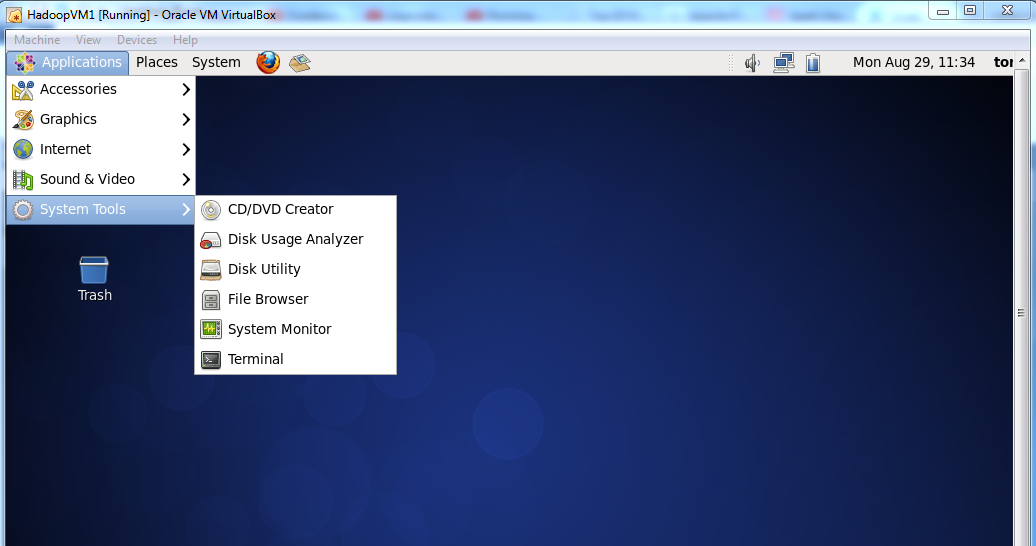
****

1. Type username and password as follows:

User name: **tom** and Password:**tomtom**

****

1. Open the terminal and.

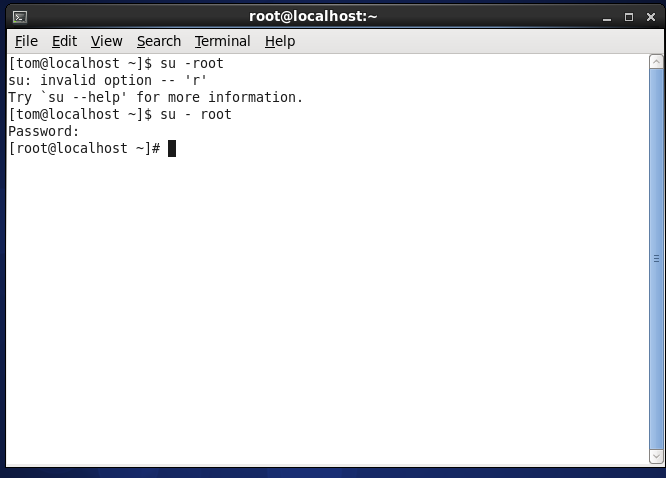
****

login to the root user to have administrator permissions with

**Command: su - root**

Type the password as follows:

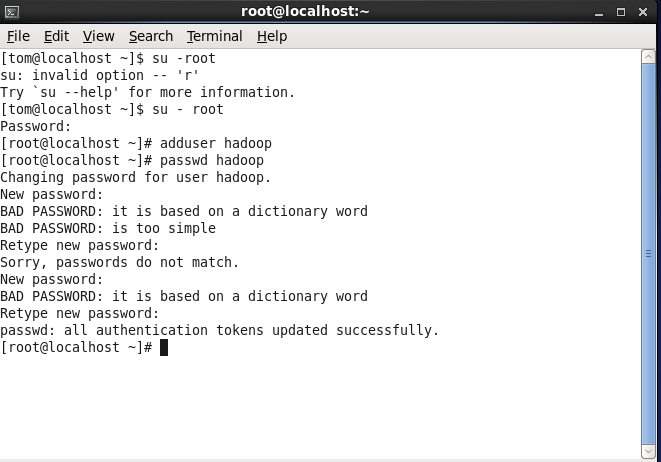
Password: **tomtom**

****

# Adding More Users in the CentOS:

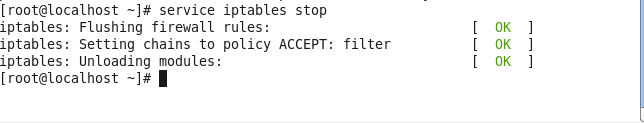
Add more users in the CentOSby

1. using the command: **adduser hadoop** and set the password of the added user by using the **command passwd hadoop1.**

****

1. Disable the firewall in the CentOS using the below command:

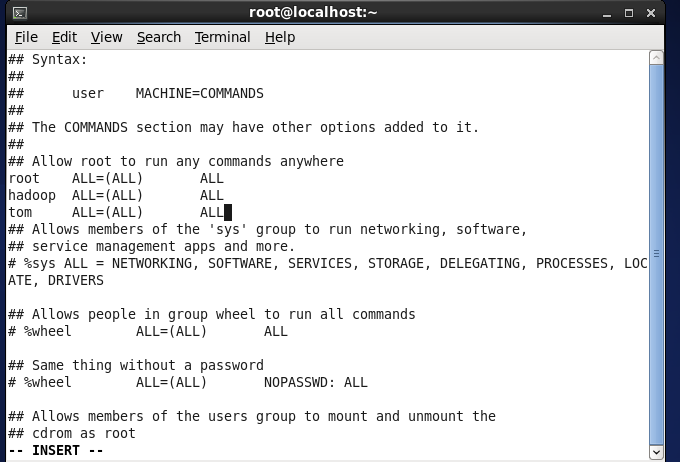
**service iptables stop**

****

1. Add the user **acadgild** into **sudoers** file to give the administrator rights to the created user.

Type the **command visudo**

1. Add the user hadoop by scrolling the cursor down as shown in the screenshot shown below:
2. To type any command press I and to save and exit press :wq

****

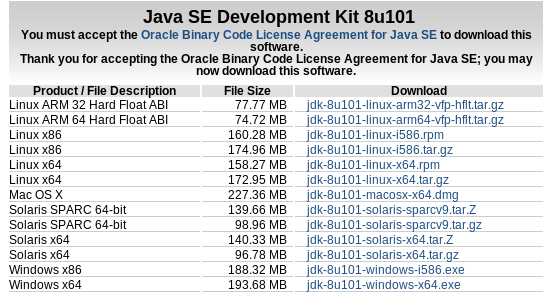
**Reboot the machine and then login to the created user.**

1. Use the below link to download **jdk** in the VM using the browser present in the centos.

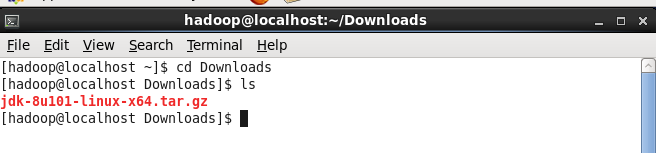
**Link:** [**http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html**](http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html)

On clicking the above link, a screen prompts for selecting the required version

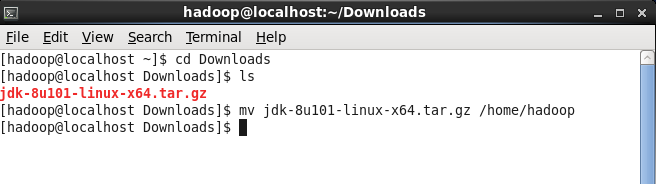
1. Select the option shown below linux-x-64 tar.gz file
2. On clicking the above option,download will start and get saved in Downloads folder.

****

Check the list by changing directory to the download folder and typing ls command

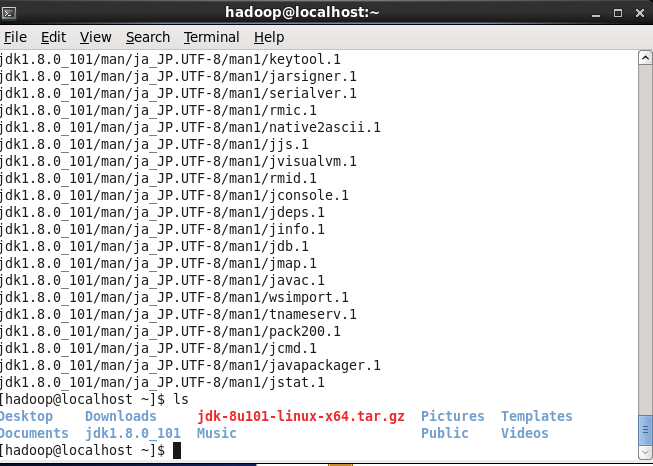
****

1. Move the above file into /home/hadoop directory using the **mv** command and then switch the directory to **/home/hadoop** by typing the command **cd**

****

****

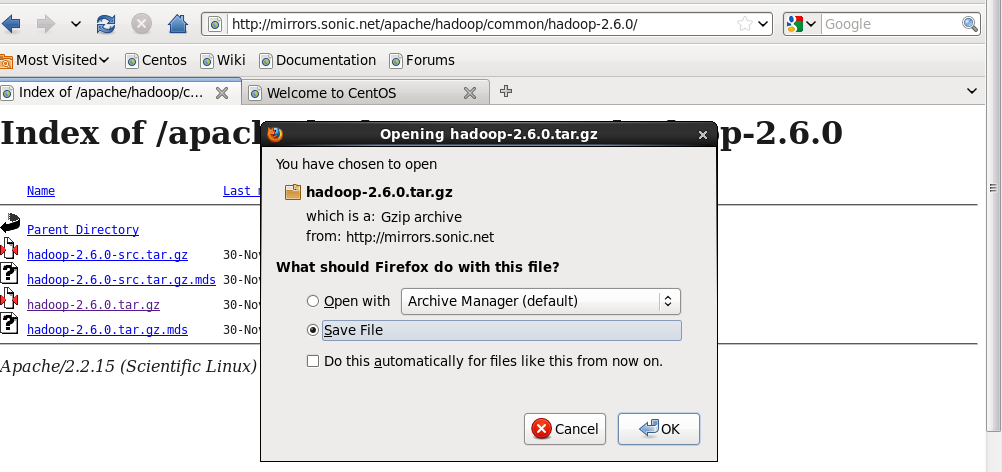
1. Unzip the jdk and extract the java file by using the command tar –xvf jdk-8u101-linux-x64.tar.gz
2. Enter the command **ls** to see the extracted jdk in the same folder **/home/hadoop**

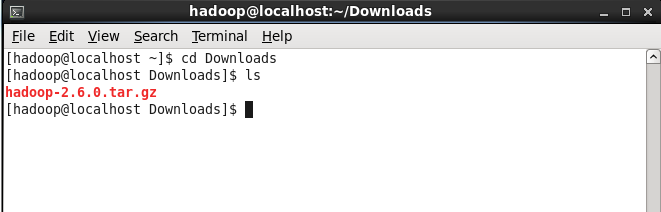
****

1. Download the hadoop file using the below link

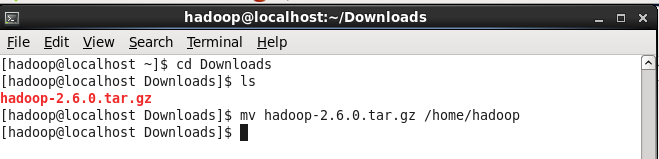
**Link:** [**http://mirrors.sonic.net/apache/hadoop/common/hadoop-2.6.0/**](http://mirrors.sonic.net/apache/hadoop/common/hadoop-2.6.0/)

On clicking the above link the below screen will prompt to select a file.

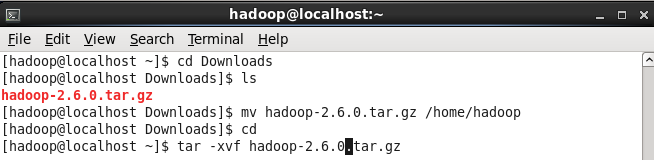
****

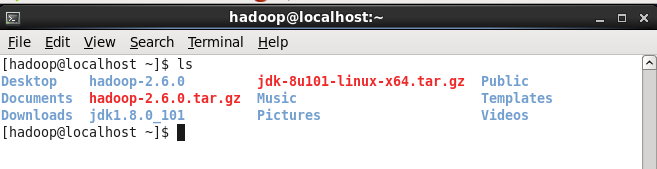
****

1. and then copy the file from Downloads folder to **/home/hadoop** directory using **mv command** as shown below

****

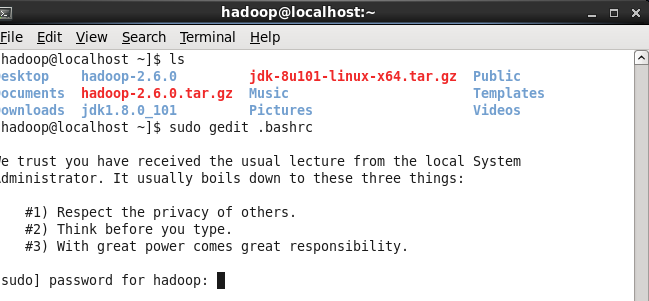
1. untar the downloaded hadoop file using the below command,refer the below screenshot.

****

****

1. Update the .**bashrc** file with required environment variables including Java and hadoop path.

Type the command **sudo gedit .bashrc** from home directory **/home/hadoop.**

****

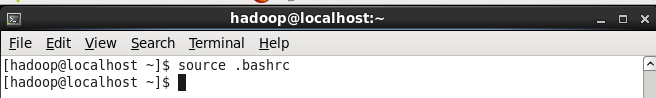
1. Update the path of system

export JAVA\_HOME=$HOME/jdk1.8.0\_101

export HADOOP\_HOME=$HOME/hadoop-2.6.0

export PATH=$JAVA\_HOME/bin:$HADOOP\_HOME/bin:$PATH

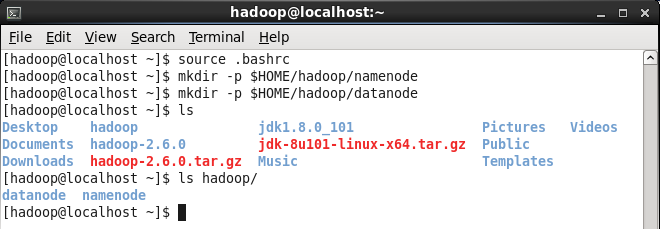
1. Type source.bashrc to make changes available.

****

1. Create two directories to store NameNode metadata and DataNode blocks as shown below:

**mkdir -p $HOME/hadoop/namenode**

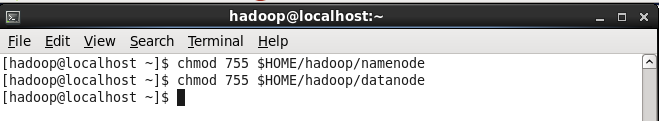
**mkdir -p $HOME/hadoop/datanode**

****

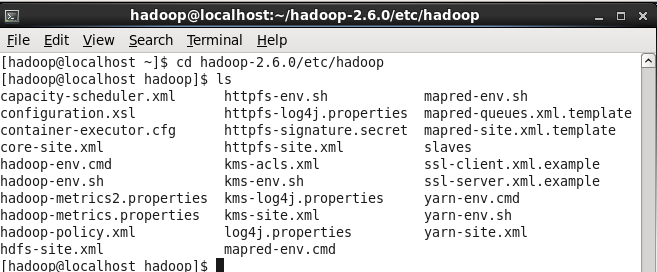
1. Change the permission of directory to 755

**chmod 755 $HOME/hadoop/namenode**

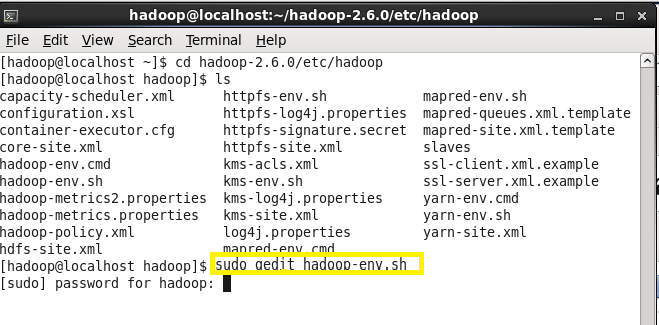
**chmod 755 $HOME/hadoop/datanode**

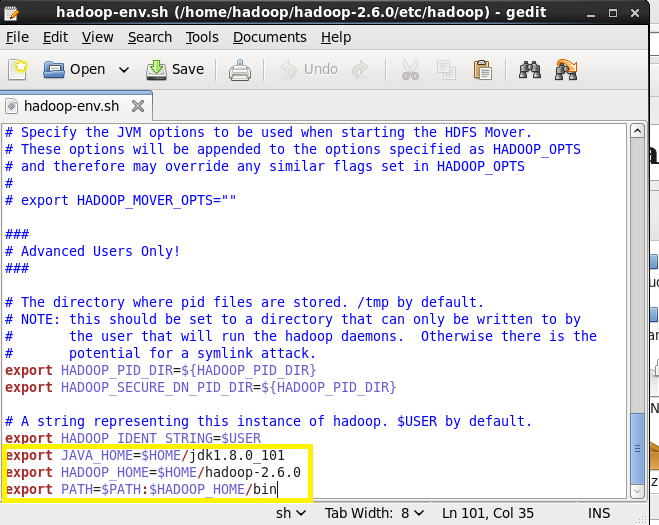
****

1. Change the directory to the location where hadoop is installed i.e hadoop-2.6.0/etc/bin

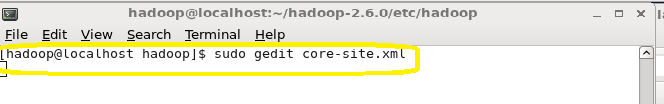
****

1. Open **hadoop-env.sh** and add the java home(path) and hadoop home(path) in it.

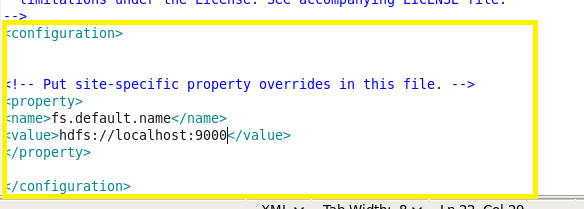
****

****

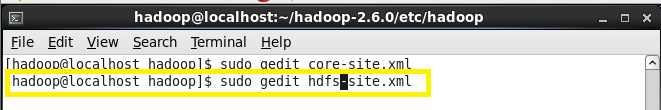
1. Open **Core-site.xml** using the below command from the path shown in the screenshot

****

1. Add the below properties in between configuration tag of core-site.xml

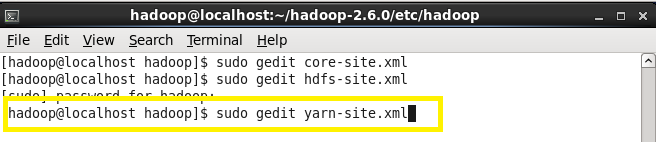
****

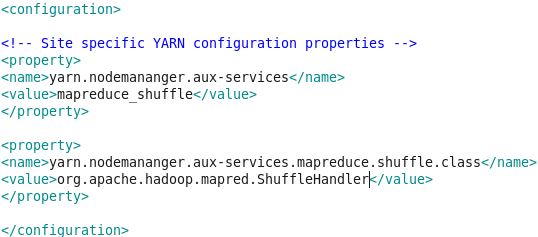
1. Open the **hdfs-site.xml** and add the following lines in between configuration tags.

****

****

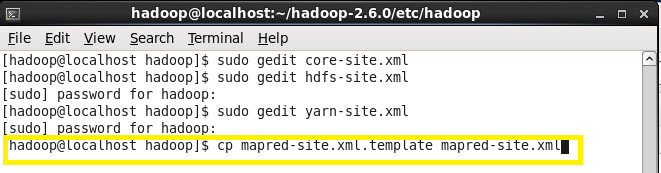
1. Open the **Yarn-site.xml** and add the following lines in between configuration tags.

****

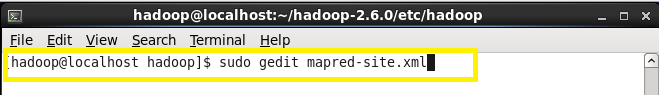
****

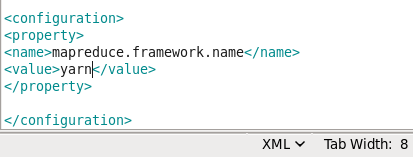
1. Copy the **mapred-site.xml** template into mapred-site.xml and then add the following properties as shown in the screenshot.

**cp mapred-site.xml.template mapred-site.xml**

****

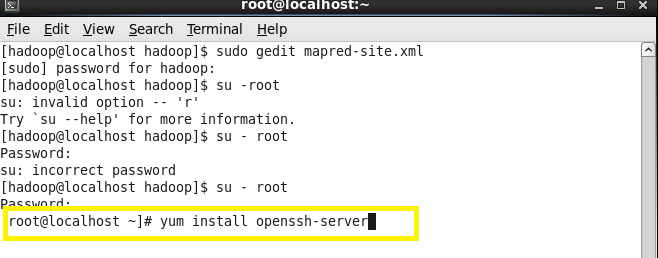
1. Open the **mapred-site.xml** and add the following lines in between configuration tags.

****

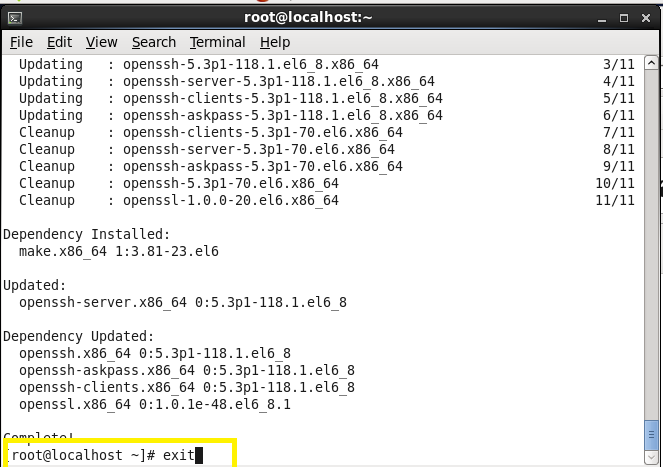
****

1. Login to the root user and then **install openssh** server in the CentOS

****

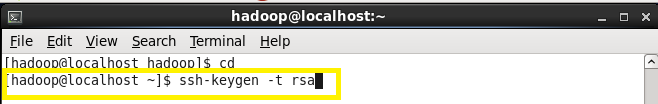
****

1. Type exit to come out of root directory

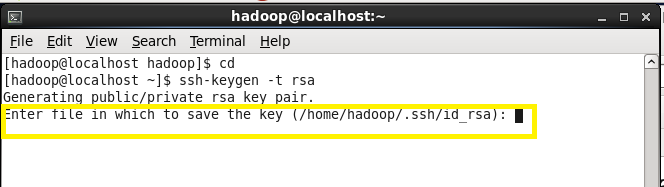
****

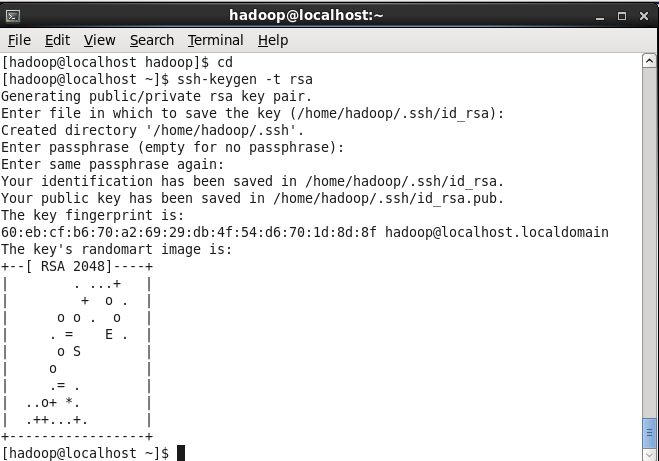
1. Generate ssh key for hadoop user.

**Command: ssh-keygen -t rsa**

****

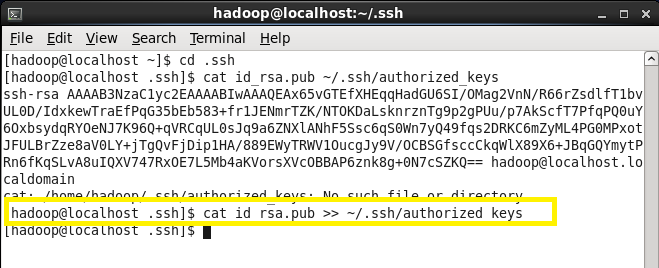
1. Hit enter

****

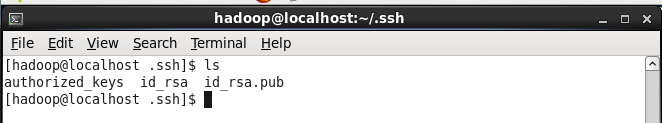
****

1. Copy the public key from .ssh directory to authorized\_keys folder.

Change the directory to **.ssh** using **command cd .ssh** and then type the below command to copy the files into the **authorized \_keys** folder.

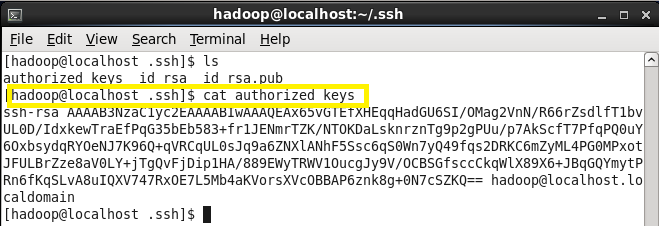
****

1. Type the command **ls** to check whether **authorized\_keys**folder has been created or not.

****

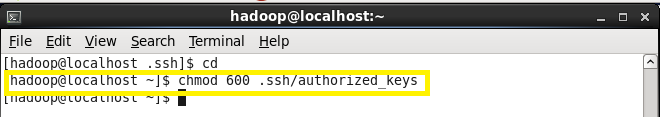
1. To ensure whether the keys have been copied, type the command

**cat authorized\_keys**

****

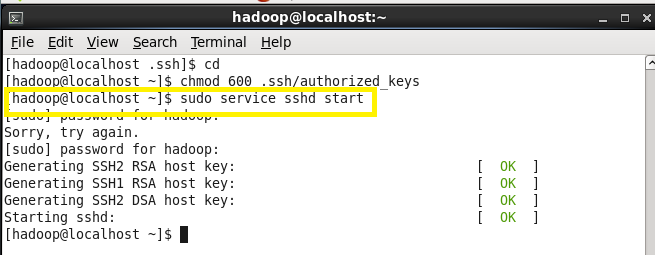
1. Change the permission of the .ssh directory.

**chmod 600 .ssh/authorized\_keys**

****

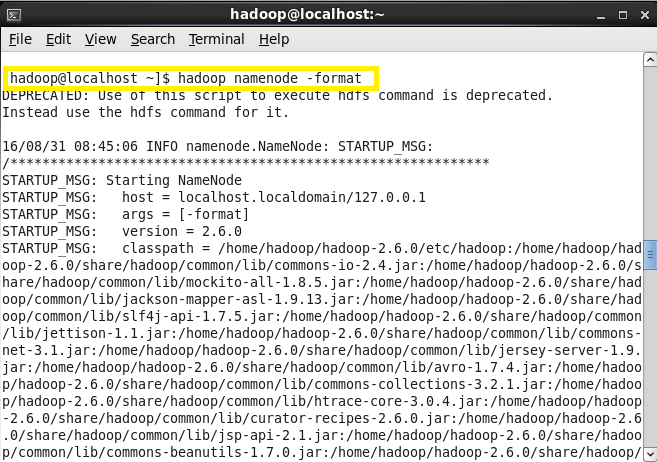
1. Restart the ssh service by typing the below command.

**Command: sudo service sshd start**

****

1. To start all the daemons follow the below steps:

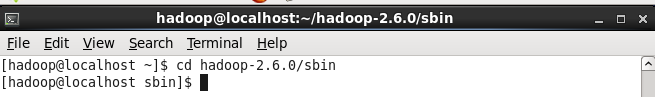
**Command: hadoop namenode -format**

****

# Starting NameNode, DataNode, ResourceManager, NodeManager and Jobhistoryserver

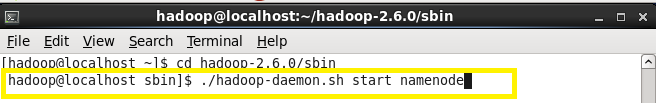
**1.** Change the directory to sbin

**Command: cd hadoop-2.6.0/sbin**

****

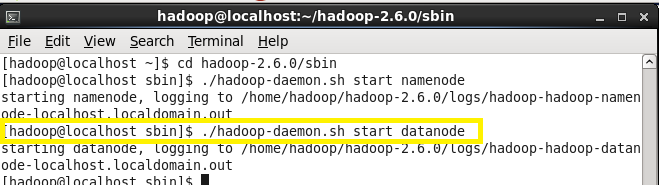
**2.** Starting namenode

**Command: ./hadoop-daemon.sh start namenode**

****

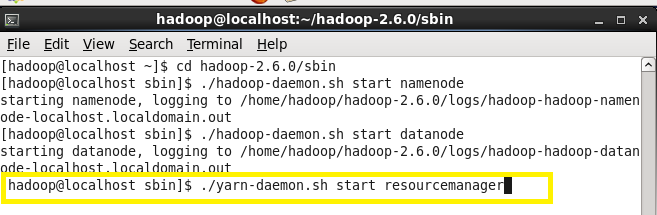
## 3. Starting DataNode

**Command: ./hadoop-daemon.sh start datanode**

****

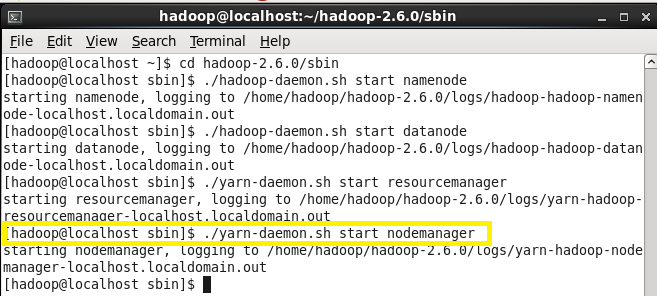
## 4. Starting ResourceManager

**Command: ./yarn-daemon.sh start resourcemanager**

****

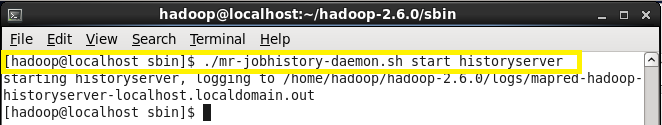
## 5. Starting NodeManager

**Command: ./yarn-daemon.sh start nodemanager**

****

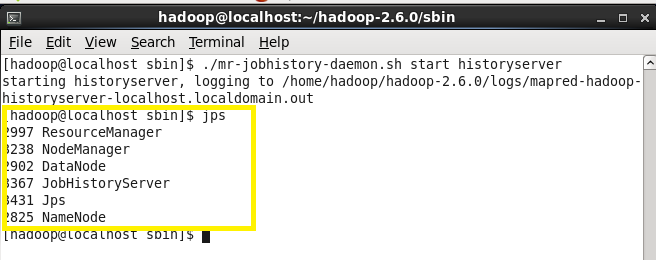
## 6. Starting Job historyserver

**Command: ./mr-jobhistory-daemon.sh start historyserver**

****

# Check Health of Daemons

Check if all the daemons have started or not by typing the command **jps**

****