Hadoop Installation Step by Step Procedure

1. Prerequisites:

Install Linux 14.04 with jdk8 in VMware Workstation. password : 123

Java 7 (at least)

To install Java 8, run: (Already installed in the machine) **//Reference only**

```
sudo add-apt-repository ppa:webupd8team/java
sudo apt-get update
sudo apt-get install oracle-java8-installer
sudo apt-get install oracle-java8-set-default
```

2. Start Installation Procedure: Open the terminal in Linux

To check installed version.

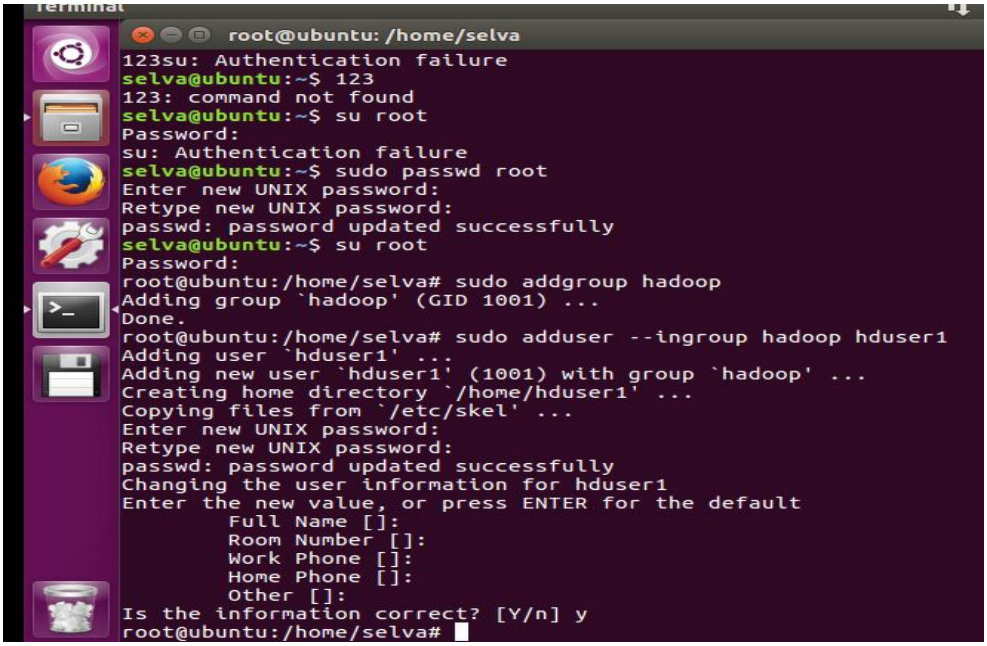
```
java -version
sudo apt-get update
```

3. Create hadoop group and new user:

You can add user only in root user. So to change root password and add user

```
sudo passwd root
(give new password for the root user) // Do remember the password given
su root
sudo addgroup hadoop
sudo adduser hduser1
sudo adduser --ingroup hadoop hduser1
sudo adduser hduser1 sudo // for giving sudo rights to hduser1
```

Provide information you are asked for. (or simply press enter)



4. **SSH Server**

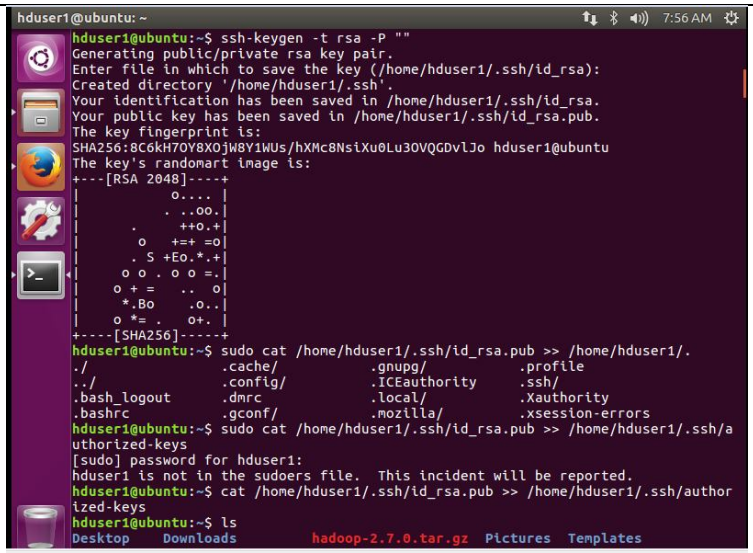
To install, run:

```
sudo apt-get install openssh-server
sudo apt-get update
```

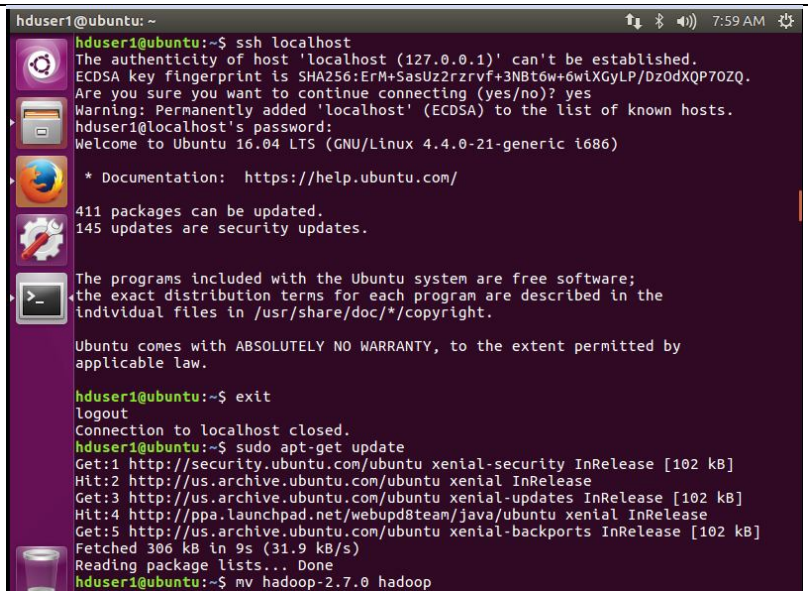
5. SSH key generation

Now go the **hduser1 l**ogin and check the java version and then Create SSH key and add it to authorized keys:

```
java -version
ssh-keygen -t rsa -P ""
```



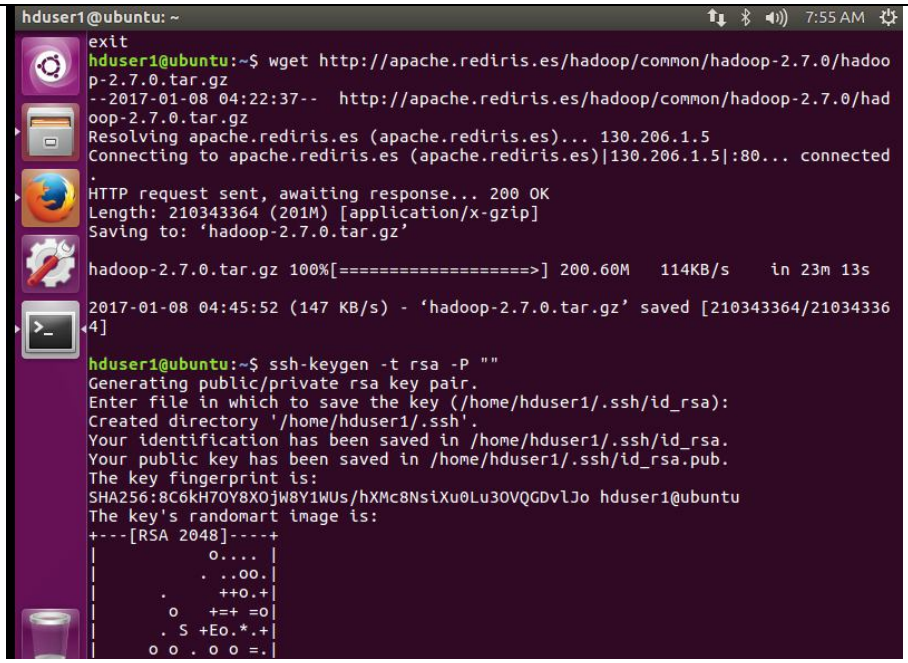

```
sudo cat /home/hduser1/.ssh/id_rsa.pub >> /home/hduser1/.ssh/authorized_keys
sudo chmod 600 /home/hduser1/.ssh/authorized_keys
ssh localhost
exit
```



6. Hadoop Installtion in hduser1

Download and unpack Apache Hadoop 2.7 :

```
sudo cp -R /home/selva/hadoop-2.7.0.tar.gz /home/hduser1/hadoop-2.7.0.tar.gz
sudo chmod 777 -R /home/hduser1/hadoop-2.7.0.tar.gz
sudo tar -xzvf hadoop-2.7.0.tar.gz
sudo mv /home/hduser1/hadoop-2.7.0 /home/hduser1/hadoop
sudo chmod 777 -R /home/hduser1/hadoop
```



7. Hadoop Configuration:

Check where your Java is installed:

```
readlink -f /usr/bin/java
```

If you get like this /usr/lib/jvm/java-8-oracle/jre/bin/java,

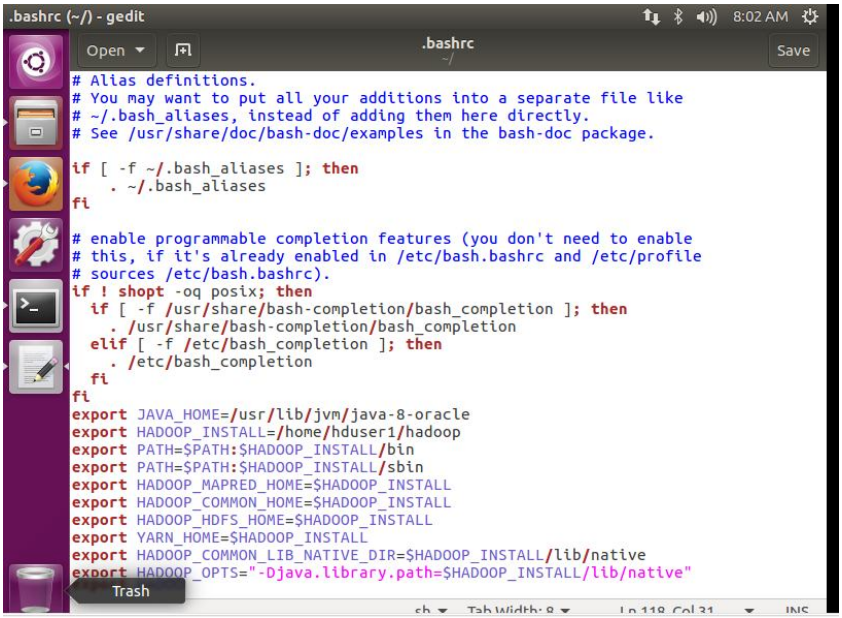
/usr/lib/jvm/java-8-oracle is what you should used for JAVA_HOME.

a) Add to ~/.bashrc file:

```
gedit ~/.bashrc
```

Add the following at the end of the file, save and close it

```
export JAVA_HOME=/usr/lib/jvm/java-8-oracle
export HADOOP_INSTALL=/home/hduser1/hadoop
export PATH=$PATH:$HADOOP_INSTALL/bin
export PATH=$PATH:$HADOOP_INSTALL/sbin
export HADOOP_MAPRED_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_HOME=$HADOOP_INSTALL
export HADOOP_HDFS_HOME=$HADOOP_INSTALL
export YARN_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_INSTALL/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_INSTALL/lib/native"
```



Reload ~/.bashrc file:

```
source ~/.bashrc
```

b)Modify JAVA_HOME in /home/hduser1/hadoop/etc/hadoop/hadoop-env.sh:

```
export JAVA_HOME=/usr/lib/jvm/java-8-oracle
```

c) Modify /home/hduser1/hadoop/etc/hadoop/core-site.xml to have something like:

```
<configuration>
  <property>
  <name>fs.default.name</name>
  <value>hdfs://localhost:9000</value>
 </property>
</configuration>
```

d) Modify /home/hduser1/hadoop/etc/hadoop/yarn-site.xmlto have something like:

```
<configuration>
 ...
 <property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
 </property>
</configuration>
```

e) Create /home/hduser1/hadoop/etc/hadoop/mapred-site.xml from template:

```
sudo cp /home/hduser1/hadoop/etc/hadoop/mapred-site.xml.template
        /home/hduser1/hadoop/etc/hadoop/mapred-site.xml
sudo chmod 777 -R /home/hduser1/hadoop/etc/hadoop/mapred-site.xml
```

f) Modify /usr/local/lib/hadoop-2.7.0/etc/hadoop/mapred-site.xml to have something like:

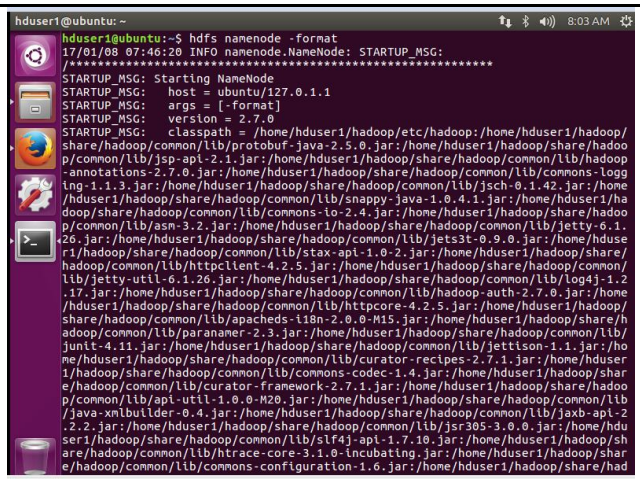
```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

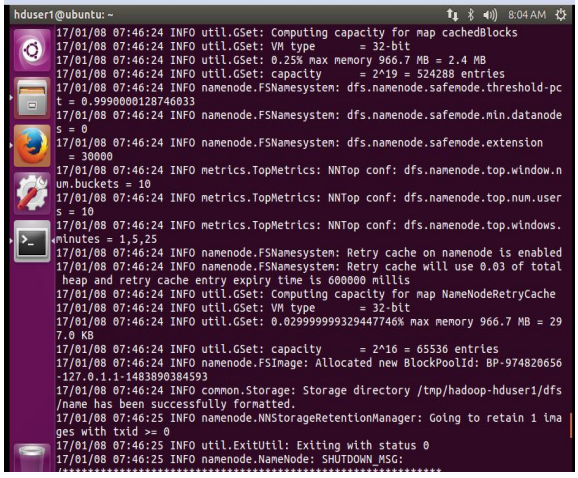
g) Modify /home/hduser1/hadoop/etc/hadoop/hdfs-site.xml to have something like:

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
</property>
</configuration>
```

Format File System

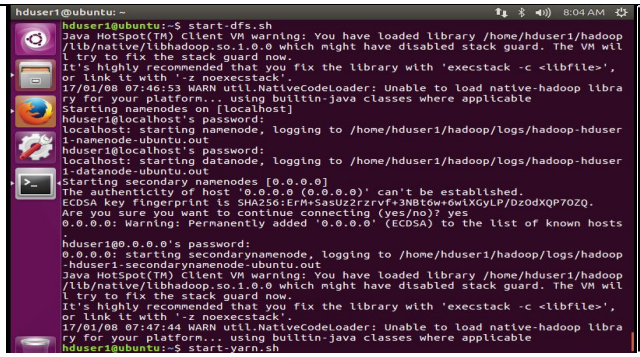
hdfs namenode -format

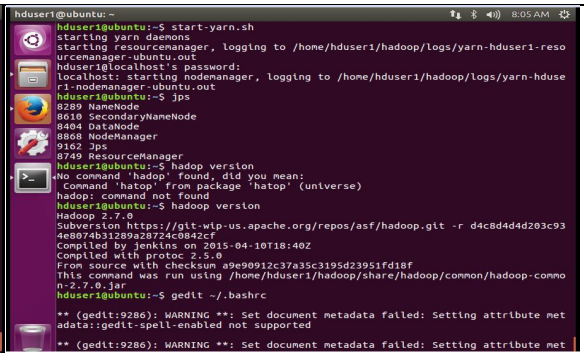




Start Hadoop:

start-all.sh





Check if everything is running:

jps

You should get something like:

Jps

NodeManager

NameNode

ResourceManager

DataNode

SecondaryNameNode

Also you can check the Hadoop through the browser http://127.0.0.1:50070