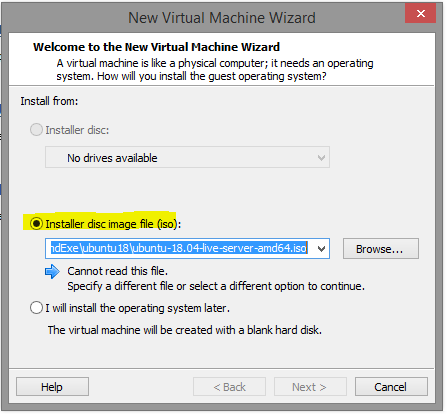
**Hadoop & Hive Installation**

**Installation of Ubuntu on VMware:**

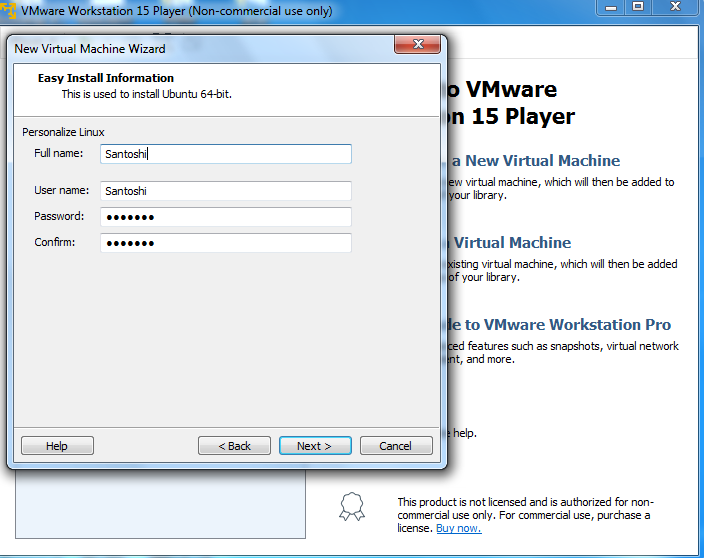
1. Using a VMware workstation to install Ubuntu on it.



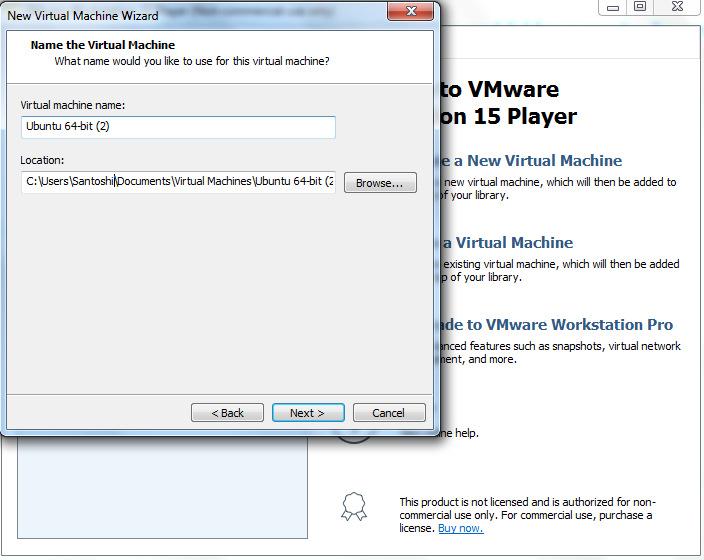
1. Downloaded the latest Ubuntu iso file from internet.
2. Pointing to the path of the downloaded Ubuntu iso image



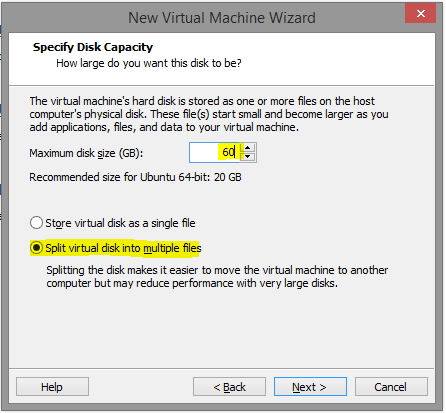
1. Entering the details to personalize



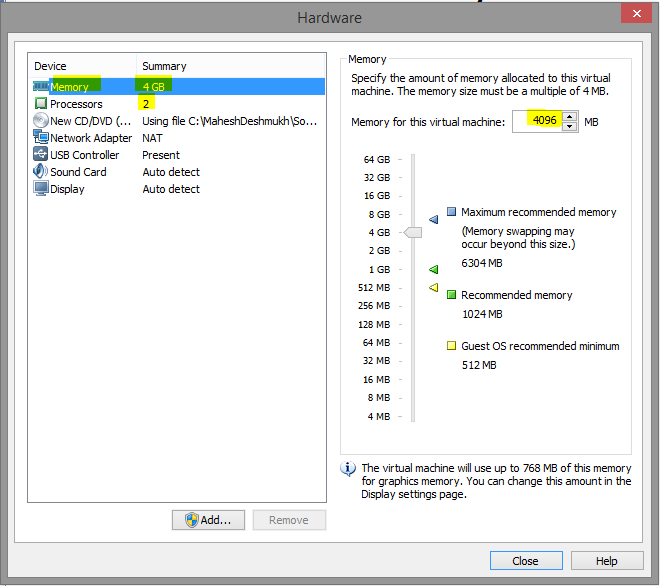
1. Providing the VM name and folder name for Binary files(.vmx)



1. Specifying the disk size.



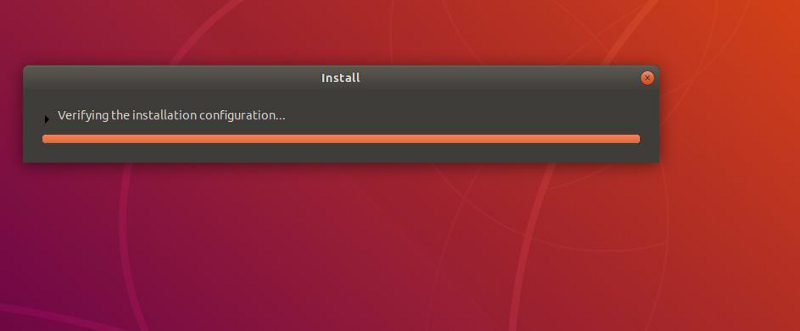
1. Customizing Hardware

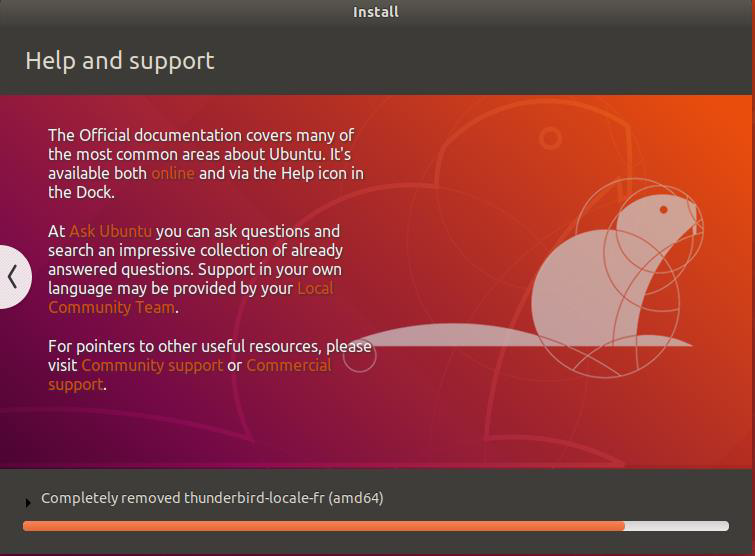


1. Start of Ubuntu installation

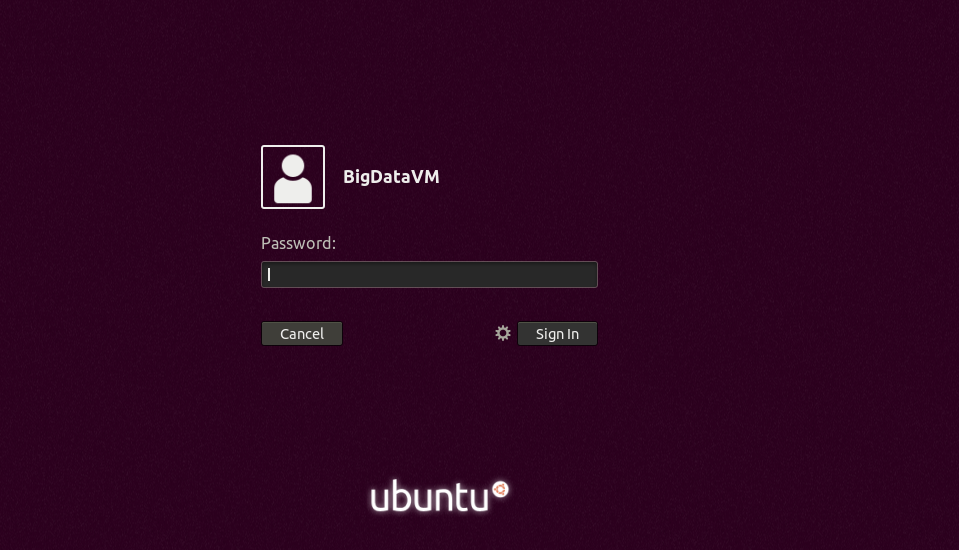


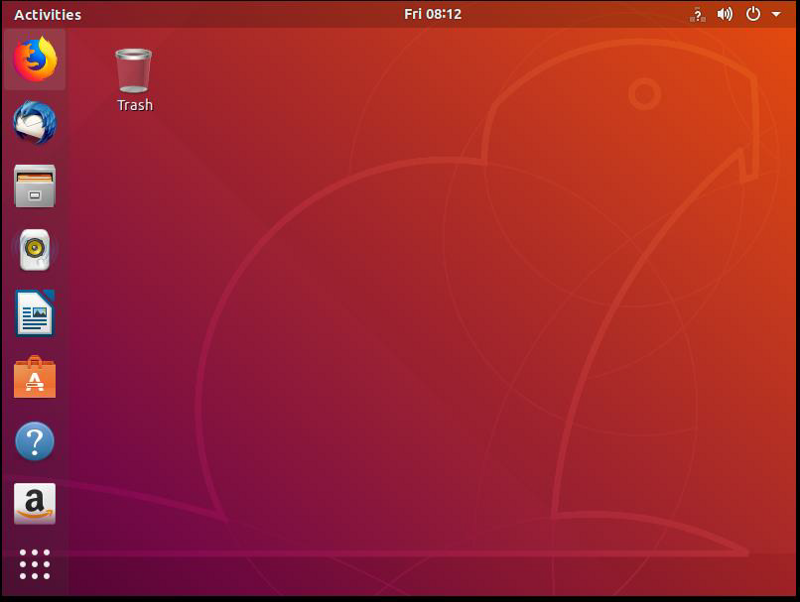




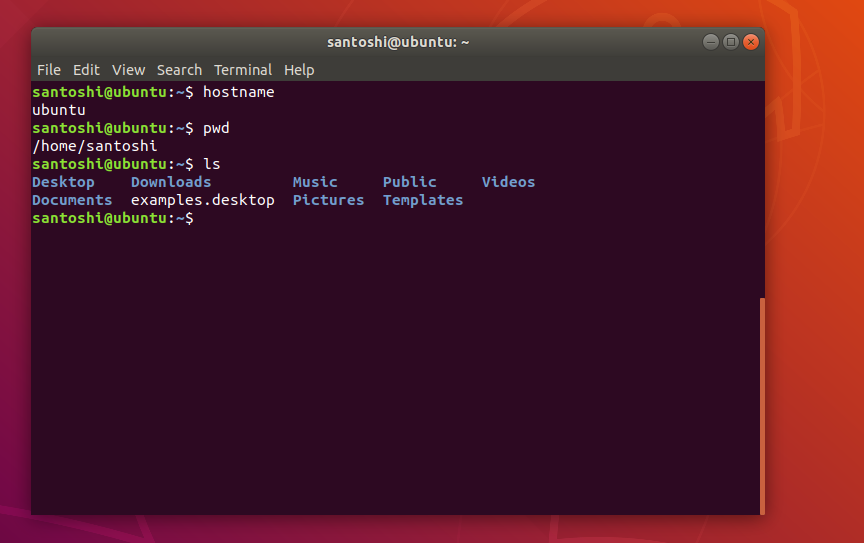


1. After successful installation, logging in to the system with the credentials specified at the beginning

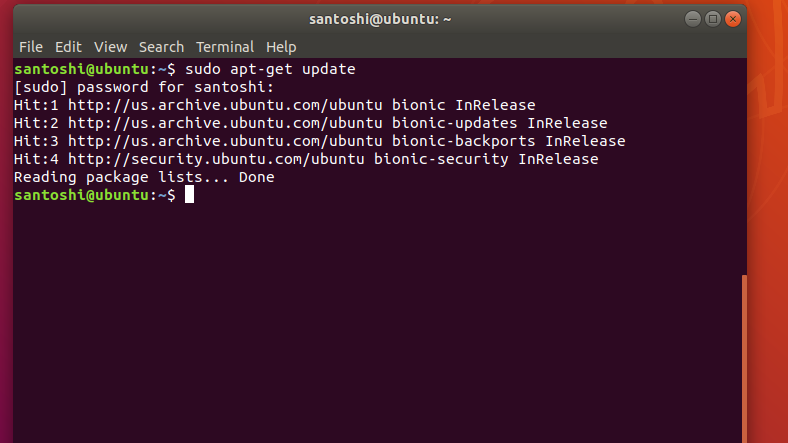


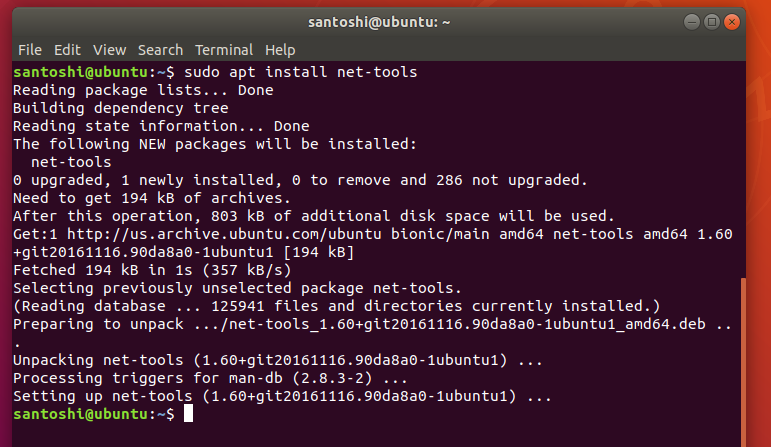


1. After successful login testing some basic commands to ensure successful installation.

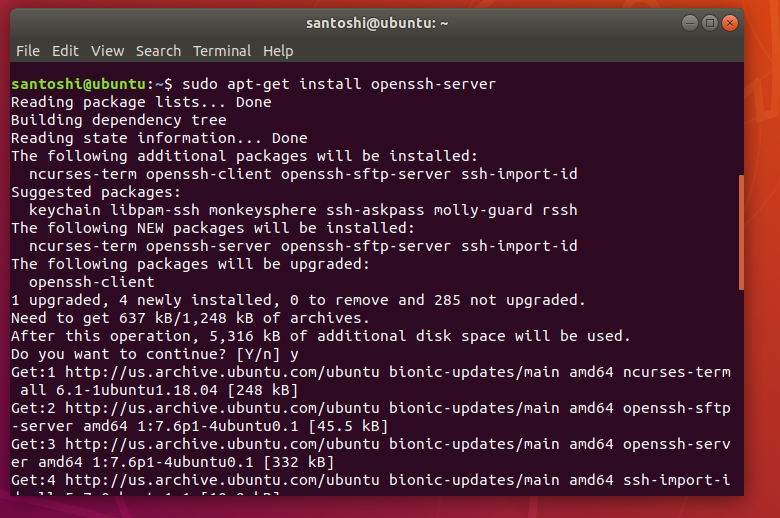


1. Installing the necessary tools.

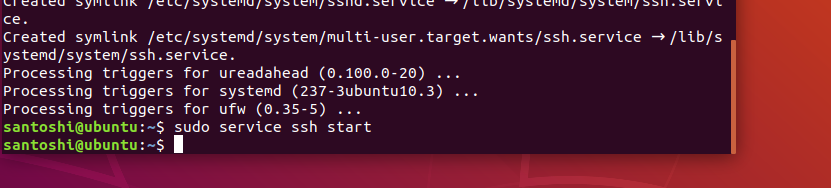




1. For accessing VM installing the necessary SSH service



1. Starting ssh — With help of this cmd we can make SSH connection through putty



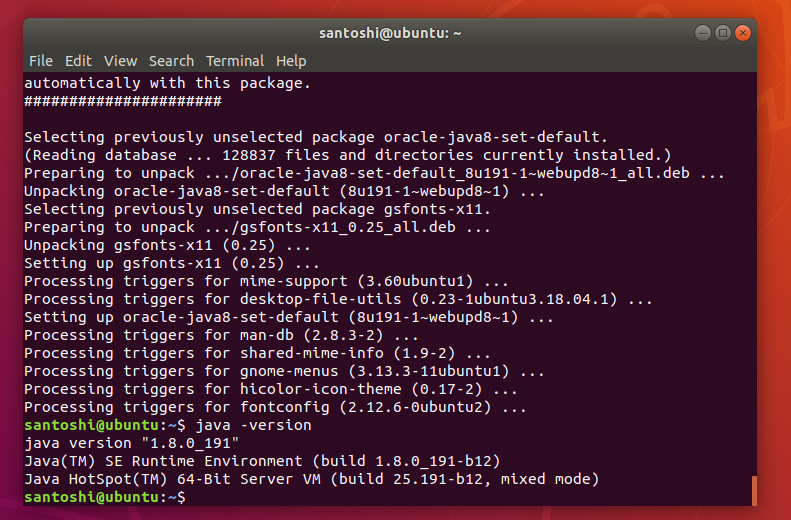
**Installation of Hadoop on Ubuntu:**

1. Installation of java

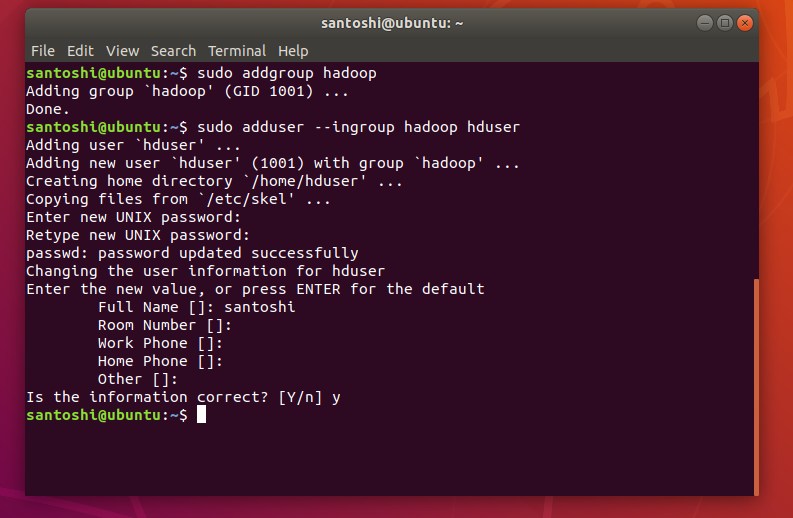
sudo add-apt-repository ppa:webupd8team/java

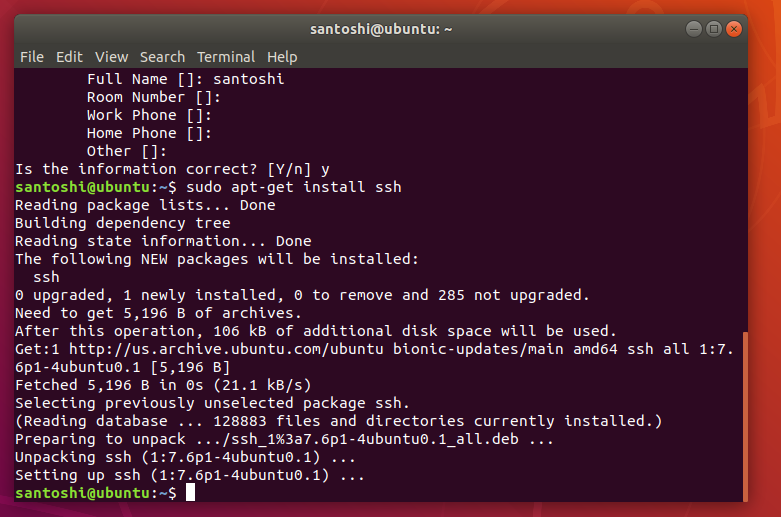
sudo apt-get update

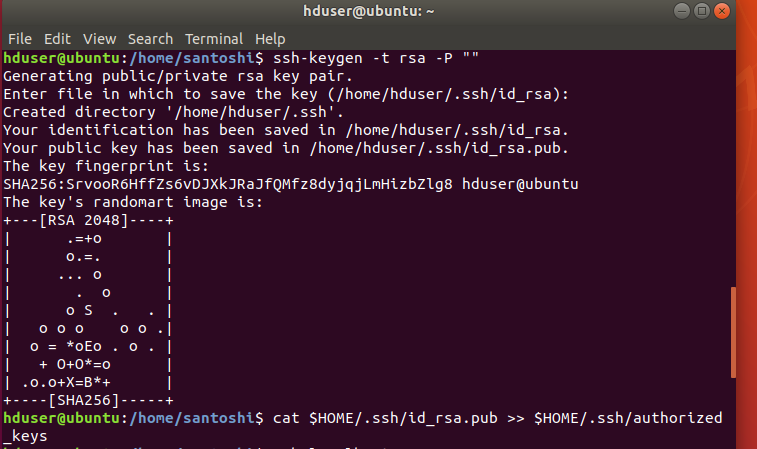
**sudo apt-get install oracle-java8-installer**

1. Checking java version
2. 

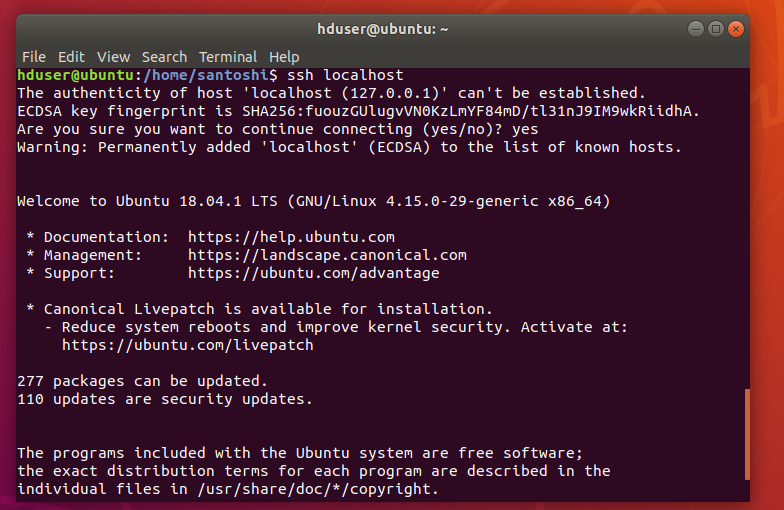
Adding Hadoop user



1. Installing SSH
2. 
3. Adding key to the list of authorized keys so that Hadoop can use ssh without prompting for a password



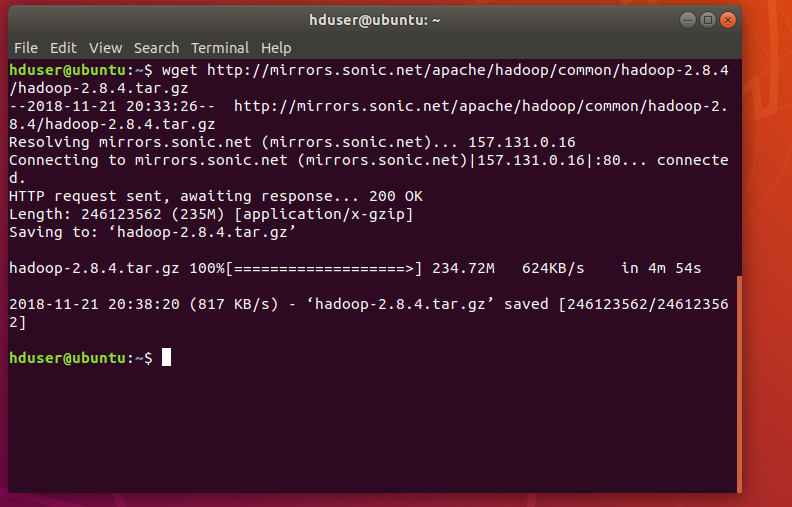
1. Checking to see if ssh is working or not



1. Downloading the latest Hadoop version

<http://mirrors.sonic.net/apache/hadoop/common/hadoop-2.8.4/hadoop-2.8.4.tar.gz>

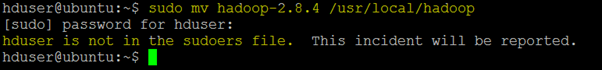
from mirror site we can get the Hadoop versions and their binaries.



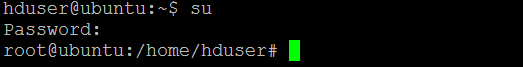
1. Extracting the tar file

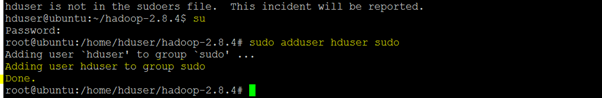
https://cdn-images-1.medium.com/max/800/1*VtzTojgOOyZwf4-8_WZ50w.png

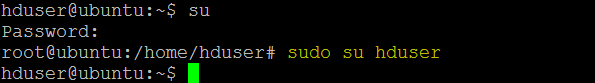
1. Moving the downloaded Hadoop dir to **/usr/local/hadoop** directory



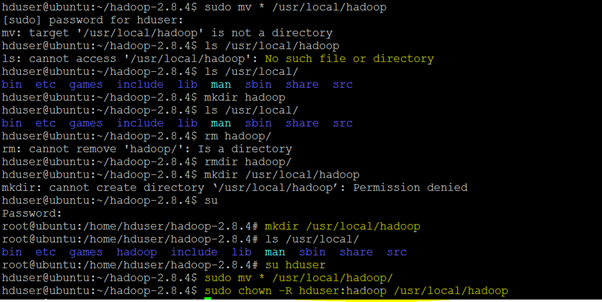
1. Adding hduser as root, and then add **hduser** to **sudo**







1. Moving all installation to /usr/local/hadoop directory



1. **Setup Configuration Files:**

The following files will have to be modified to complete the Hadoop setup:

1. ~/.bashrc

2. /usr/local/hadoop/etc/hadoop/hadoop-env.sh

3. /usr/local/hadoop/etc/hadoop/core-site.xml

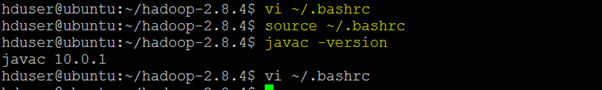
4. /usr/local/hadoop/etc/hadoop/mapred-site.xml.template

5. /usr/local/hadoop/etc/hadoop/hdfs-site.xml

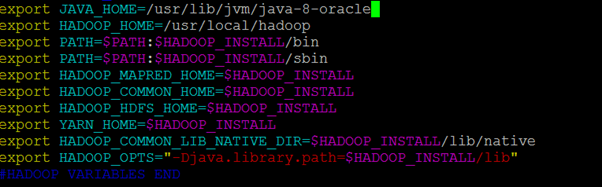
**1. ~/.bashrc:**

To find the path where Java has been installed to set the JAVA\_HOME environment variable:

https://cdn-images-1.medium.com/max/800/1*auYgnX4gPvRTFwbvyizM6g.png

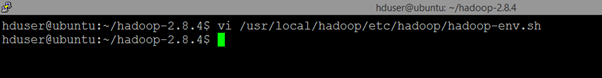


**Environment variables entered in .bashrc file.**



**2. /usr/local/hadoop/etc/hadoop/hadoop-env.sh**

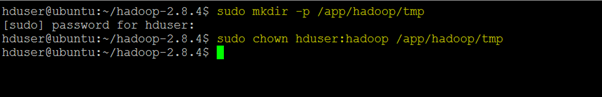
Setting **JAVA\_HOME** by modifying hadoop-env.sh file.



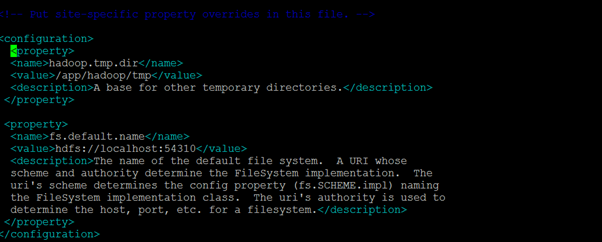
https://cdn-images-1.medium.com/max/800/1*0s89dm-mL3kd5Tv2gzhhuw.png

**3. /usr/local/hadoop/etc/hadoop/core-site.xml:**

The /usr/local/hadoop/etc/hadoop/core-site.xml file contains configuration properties that Hadoop uses when starting up.   
This file can be used to override the default settings that Hadoop starts with.



https://cdn-images-1.medium.com/max/800/1*Cs5Bg-P1YtNBWbNFwQssYA.png

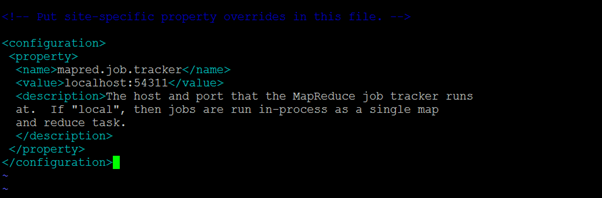


4. /usr/local/hadoop/etc/hadoop/mapred-site.xml

By default, the /usr/local/hadoop/etc/hadoop/ folder contains /usr/local/hadoop/etc/hadoop/mapred-site.xml.template file which has to be renamed/copied with the name **mapred-site.xml**:

https://cdn-images-1.medium.com/max/800/1*Y82fmMNmjqjmaHWxTfwnCA.png

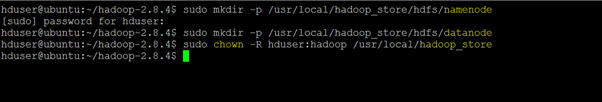
https://cdn-images-1.medium.com/max/800/1*AY5bYQrLFFHgU2p5aCD4DA.png



**5. /usr/local/hadoop/etc/hadoop/hdfs-site.xml**

The /usr/local/hadoop/etc/hadoop/hdfs-site.xml file needs to be configured for each host in the cluster that is being used.   
 It is used to specify the directories which will be used as the **namenode** and the **datanode** on that host.

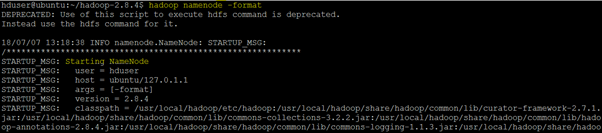
Before editing this file, creating two directories which will contain the namenode and the datanode for this Hadoop installation. 



https://cdn-images-1.medium.com/max/800/1*4qsaYorMs7dQNmr_XkwmDg.png

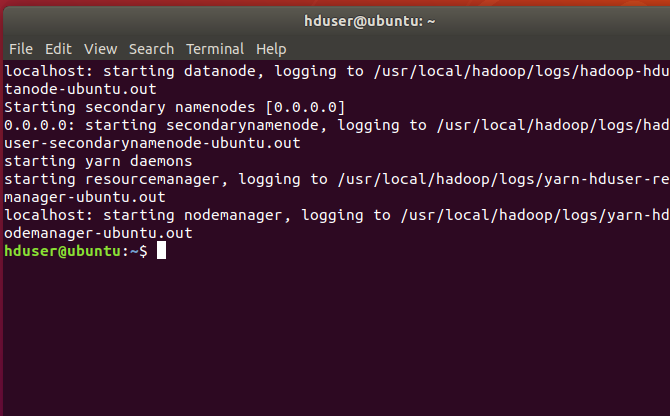


1. **Formatting the New Hadoop Filesystem**

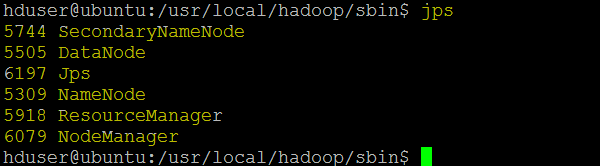


1. **Starting Hadoop**

**Start Hadoop by executing the start-all.sh command from Hadoop/sbin folder**



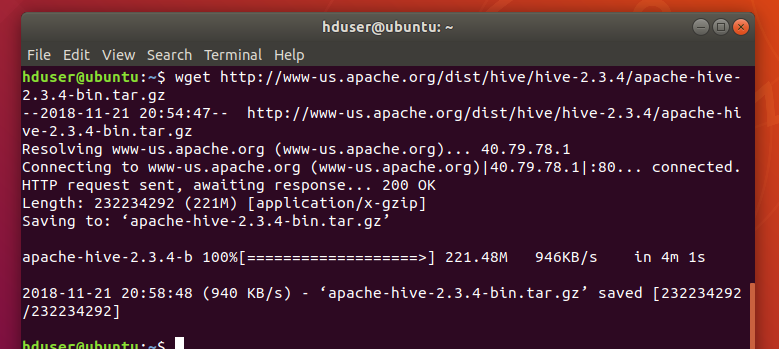
1. Checking the services running



**Installation of Hive on Ubuntu:**

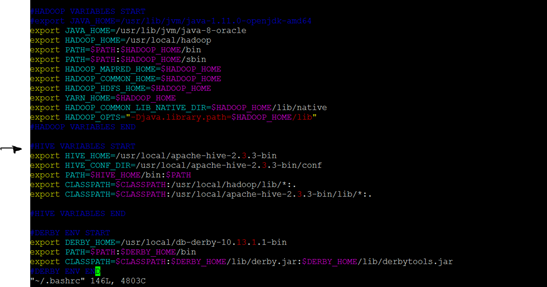
1. Downloading the latest version of hive

wget <http://www-us.apache.org/dist/hive/hive-2.3.4/apache-hive-2.3.4-bin.tar.gz>

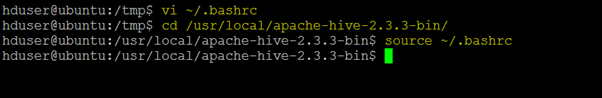


1. Setting the hive environment variables in bashrc

Openning ~/.bashrc file and setting the environment variable **HIVE\_HOME** to point to the installation directory and **PATH**:

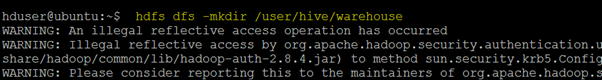


1. Running the update



1. **Creating Hive warehouse directory**

https://cdn-images-1.medium.com/max/800/1*KqIbiGTfdOewvIYDIdlqlQ.png



1. **Configuring Hive**

**Copy hive-env template to hive-en.sh**

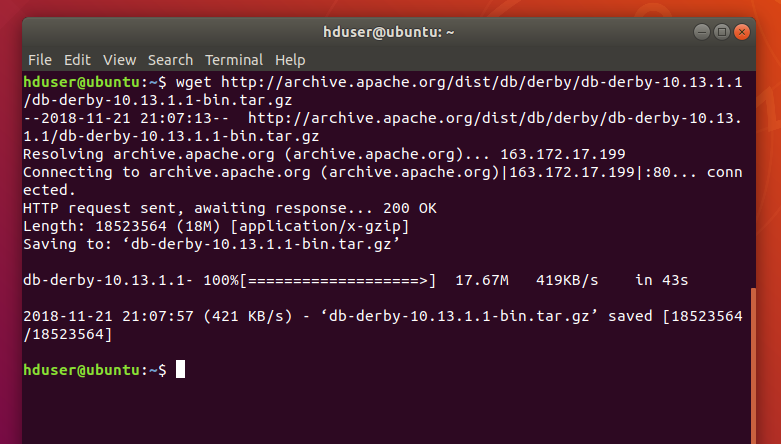
https://cdn-images-1.medium.com/max/800/1*8JMfFAqI3eustHkgjYeuxg.png

Editing the hive-env.sh file by appending the following line:

**export HADOOP\_HOME=/usr/local/Hadoop**

1. Installing Apache Derby

**wget**[**http://archive.apache.org/dist/db/derby/db-derby-10.13.1.1/db-derby-10.13.1.1-bin.tar.gz**](http://archive.apache.org/dist/db/derby/db-derby-10.13.1.1/db-derby-10.13.1.1-bin.tar.gz)

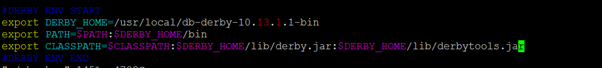
1. 
2. Configuring Derby

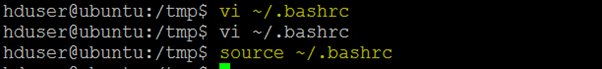
Setting up the Derby environment by appending the following lines to ~/.bashrc file:

**export DERBY\_HOME=/usr/local/db-derby-10.13.1.1-bin**

**export PATH=$PATH:$DERBY\_HOME/bin**

**export CLASSPATH=$CLASSPATH:$DERBY\_HOME/lib/derby.jar:$DERBY\_HOME/lib/derbytools.jar**





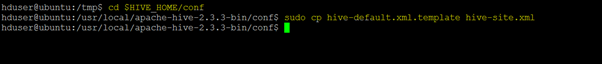
Creating a directory named data in $**DERBY\_HOME**directory to store Metastore data.

https://cdn-images-1.medium.com/max/800/1*Ar1R2Jebdxarlfa0LAvbWg.png

1. **Configuring Hive Metastore :**

Editing the **hive-site.xml file**, which is in the $HIVE\_HOME/conf directory.

copying the template file using the following command:



<property>

<name>javax.jdo.option.ConnectionURL</name>

<value>jdbc:derby:;databaseName=metastore\_db;create=true</value>

<description>

JDBC connect string for a JDBC metastore.

To use SSL to encrypt/authenticate the connection, provide database-specific SSL flag in the connection URL.

For example, jdbc:postgresql://myhost/db?ssl=true for postgres database.  
</description>  
</property>

1. Creating a file named jpox.properties

Adding the following lines

javax.jdo.PersistenceManagerFactoryClass =

org.jpox.PersistenceManagerFactoryImpl  
org.jpox.autoCreateSchema = false  
org.jpox.validateTables = false  
org.jpox.validateColumns = false  
org.jpox.validateConstraints = false  
org.jpox.storeManagerType = rdbms  
org.jpox.autoCreateSchema = true  
org.jpox.autoStartMechanismMode = checked  
org.jpox.transactionIsolation = read\_committed  
javax.jdo.option.DetachAllOnCommit = true  
javax.jdo.option.NontransactionalRead = true

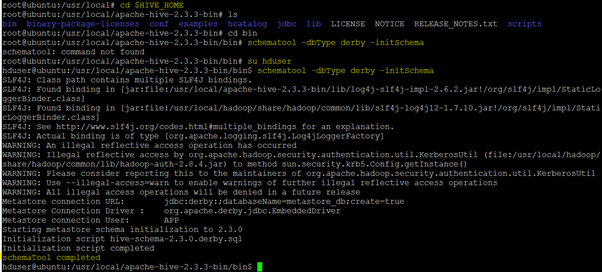
javax.jdo.option.ConnectionDriverName = org.apache.derby.jdbc.ClientDriver

javax.jdo.option.ConnectionURL = jdbc:derby://hadoop1:1527/metastore\_db;create = true

javax.jdo.option.ConnectionUserName = APP

javax.jdo.option.ConnectionPassword = mine

1. **Metastore schema initialization**



1. **Verifying Hive Installation by running Hive CLI**

To use the Hive command line interface (CLI) from the shell, issue bin/hive command to verify Hive

https://cdn-images-1.medium.com/max/800/1*Kbc7X9Q4JTeW4lN8JSp3Xw.png

