# How to Install Hadoop on Linux

## Installing Hadoop

Installing and getting Hadoop is quite complex. I will describe the required steps by step for install Hadoop on linux. I install as root

### 1. Install Java

Hadoop requires Java to be installed. Execute below comments

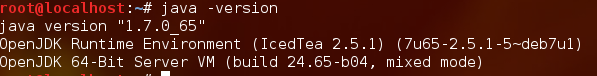
apt-get update

apt-get install default-jdk

These commands will update the package Java .

java -version

If Java has been installed, this command show java version that installed on your system



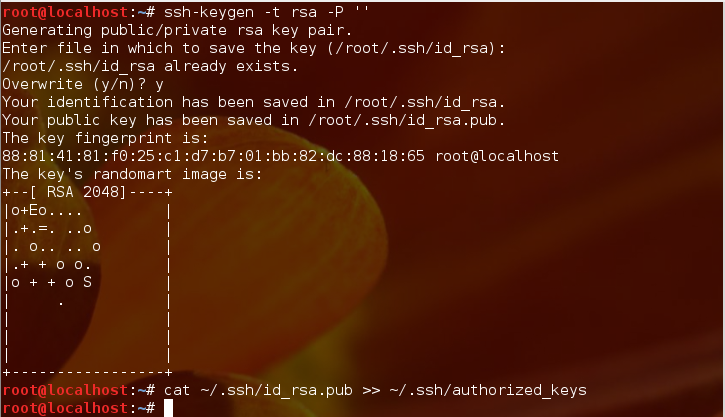
### 2. Install and Create SSH

Hadoop uses SSH to access its nodes. So SSH should be installed

ssh-keygen -t rsa -P ''

cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys

Executing this above command. System will be asked for a filename, press enter and go on. Hadoop use SSH Without prompting for a password.



### 3. Download and Install Hadoop

First let's download Hadoop from one of the mirrors using the following command:

wget http://ftp.itu.edu.tr/Mirror/Apache/hadoop/common/hadoop-2.5.1/

After downloading the Hadoop , execute the below command to extract it:

tar xfz hadoop-2.5.1.tar.gz

This command will extract all the files in this package in a directory named

Hadoop-2.5.1 and we have to move this file under the /usr/local file

Execute above command to move hadoop in /usr/local/hadoop

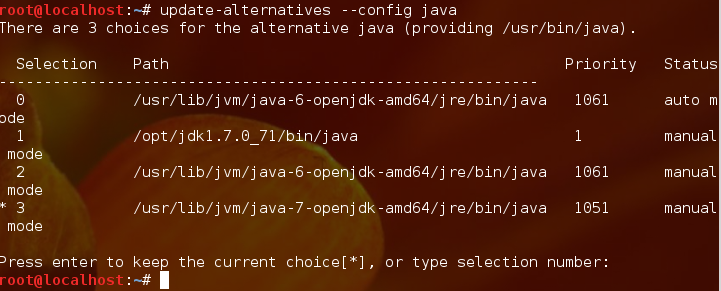
mv hadoop-2.3.0 /usr/local/hadoop

4.Editing ~/.bashrc

Firstly we have to know where java installed. Execute above command

update-alternatives --config java

This will display something like the following, Or maybe you have one java.



The complete path displayed by this command is:

/usr/lib/jvm/java-7-openjdk-amd64

Use  nano  using the following command:

nano ~/.bashrc

This will open the .bashrc file in a text editor. Go to the end of the file and paste/type the following content in it:

#HADOOP VARIABLES START

export JAVA\_HOME=/usr/lib/jvm/java-7-openjdk-amd64

export HADOOP\_INSTALL=/usr/local/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

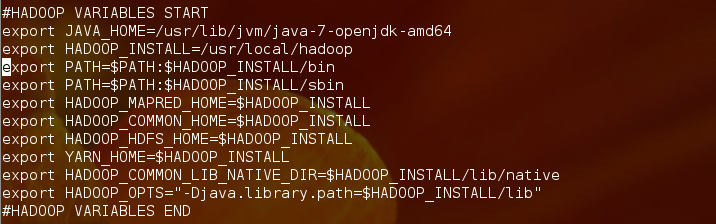
export YARN\_HOME=$HADOOP\_INSTALL

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_INSTALL/lib/native

export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_INSTALL/lib"

#HADOOP VARIABLES END

The end of the .bashrc file should look something like this:



### 5. Editing /usr/local/hadoop/etc/hadoop/hadoop-env.sh

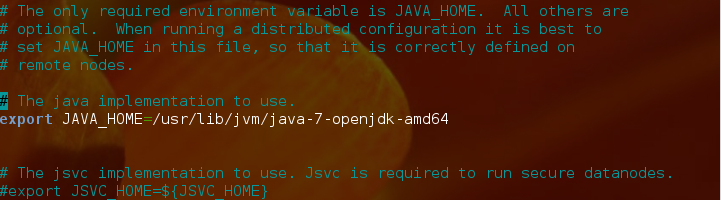
Open the /usr/local/hadoop/etc/hadoop/hadoop-env.sh file with nano using the following command:

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

In this file, locate the line that exports the JAVA\_HOME variable. Change this line to the following:

export JAVA\_HOME=/usr/lib/jvm/java-7-openjdk-amd64

The hadoop-env.sh file should look something like this:



Save and quit .

### 6. Editing /usr/local/hadoop/etc/hadoop/core-site.xml

The /usr/local/hadoop/etc/hadoop/core-site.xml . nano using the following command:

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

In this file, enter the following content in between the <configuration></configuration> tag:

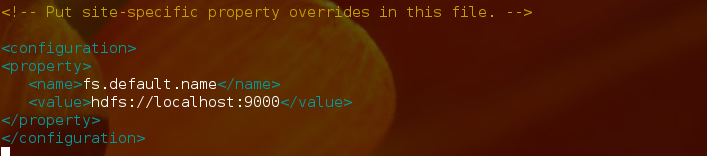
<property>

<name>fs.default.name</name>

<value>hdfs://localhost:9000</value>

</property>

The core-site.xml file should look something like this:



Save and close this file.

### 7. Editing /usr/local/hadoop/etc/hadoop/yarn-site.xml

The /usr/local/hadoop/etc/hadoop/yarn-site.xml file contains configuration properties that MapReduce uses when starting up.

Open this file with nano using the following command:

nano /usr/local/hadoop/etc/hadoop/yarn-site.xml

In this file, enter the following content in between the <configuration></configuration> tag:

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

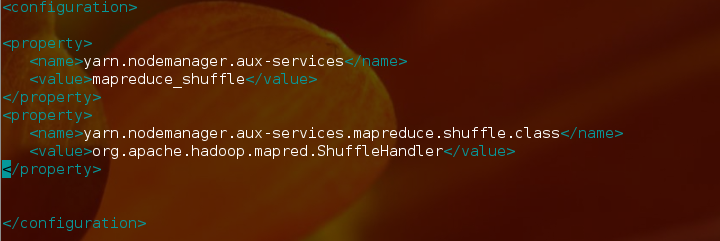
<property>

<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

The yarn-site.xml file should look something like this:



Save and close this file.

### 8. Creating and Editing /usr/local/hadoop/etc/hadoop/mapred-site.xml

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

Run following comments

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

This is done, open new file with nano using the following command:

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

In this file, enter the following content in between the <configuration></configuration> tag:

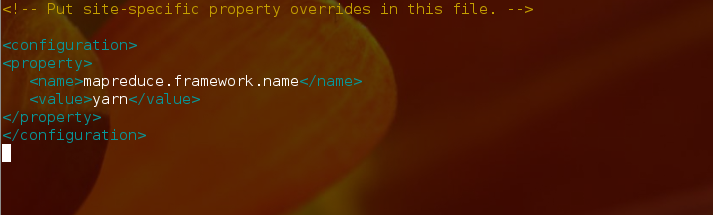
<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

The mapred-site.xml file should look something like this,



Save and close this file.

### 9. Editing /usr/local/hadoop/etc/hadoop/hdfs-site.xml

The /usr/local/hadoop/etc/hadoop/hdfs-site.xml  has to be configured the namenode and the datanode on that host.

We need to create two directories which will contain the namenode and the datanode for this Hadoop installation. This can be done using the following commands: Execute following Comments

mkdir -p /usr/local/hadoop\_store/hdfs/namenode

mkdir -p /usr/local/hadoop\_store/hdfs/datanode

Open the /usr/local/hadoop/etc/hadoop/hdfs-site.xml file with nano using the run following command:

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

In this file, enter the following content in between the <configuration></configuration> tag:

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:/usr/local/hadoop\_store/hdfs/namenode</value>

</property>

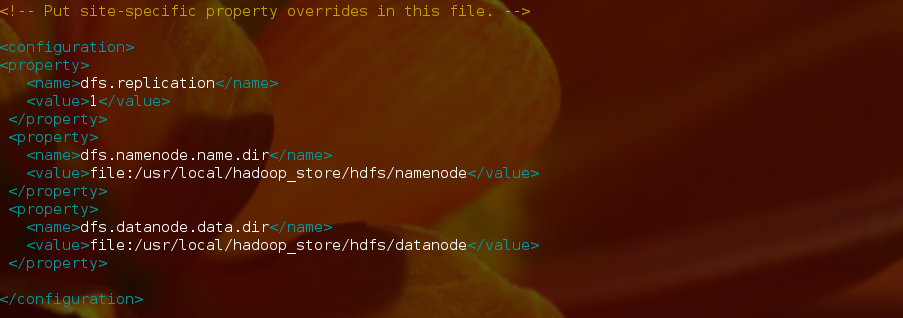
<property>

<name>dfs.datanode.data.dir</name>

<value>file:/usr/local/hadoop\_store/hdfs/datanode</value>

</property>

The hdfs-site.xml file should look something like this:



Save and close this file.

## 10.Format the New Hadoop Filesystem

After completed this above operation Hadoop file system must be format

Run following commands

hdfs namenode -format

## 11.Start Hadoop

start-all.sh

After run this comment open web browser and go to localhost:50070 or localhost:50030 and look localhost:8088

Executing above comment and look which operation is running

jps

Sholuld be look like this

## 12. Stop Hadoop

Execute following commend and stop hadoop

stop-all.sh