

# Hadoop Installation

## 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)

# Hadoop Installation

## 4. SSH Server

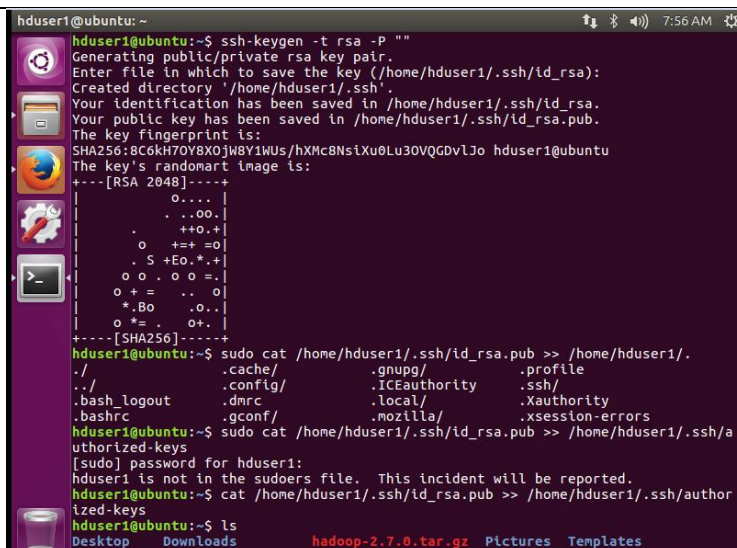
To install, run:

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

## 5. SSH key generation

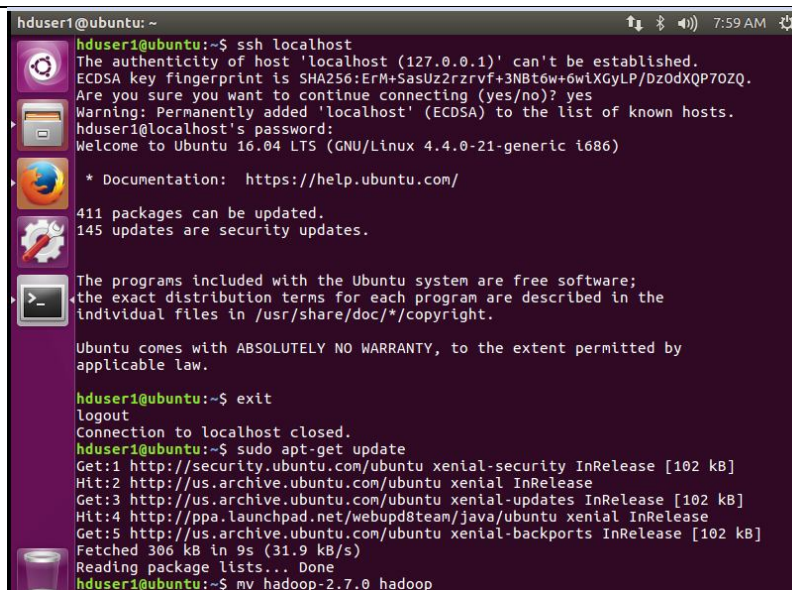
Now go the **hduser1** login and check the java version and then Create SSH key and add it to authorized keys:

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

A terminal window showing the execution of 'ssh-keygen -t rsa -P ""'. It prompts for a file to save the key, creates a directory, and shows the key fingerprint. Then, it prompts to add the key to the authorized\_keys file, which is done using 'cat' and 'echo'. The terminal also shows a list of files in the home directory.

```
hduser1@ubuntu:~$ ssh-keygen -t rsa -P ""
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hduser1/.ssh/id_rsa):
Created directory '/home/hduser1/.ssh'.
Your identification has been saved in /home/hduser1/.ssh/id_rsa.
Your public key has been saved in /home/hduser1/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:8C6kH70Y8X0JW8Y1WUs/hXM8NsIXu0Lu30VQGDv1Jo hduser1@ubuntu
The key's randomart image is:
+----[RSA 2048]-----+
|
| o....
| ..oo.
| ++o.
| o
| +--+ o=
| .S +Eo.*.+
| o o . o o =
| o + = .. o
| *.Bo .o..
| o * = . o+
+----[SHA256]-----+
hduser1@ubuntu:~$ sudo cat /home/hduser1/.ssh/id_rsa.pub >> /home/hduser1/.
./
.cache/ .gnupg/ .profile
./
.config/ .ICEauthority .ssh/
.bash_logout .dnrc .local/ .Xauthority
.bashrc .gconf/ .mozilla/ .xsession-errors
hduser1@ubuntu:~$ sudo cat /home/hduser1/.ssh/id_rsa.pub >> /home/hduser1/.ssh/a
uthorized-keys
[sudo] password for hduser1:
hduser1 is not in the sudoers file. This incident will be reported.
hduser1@ubuntu:~$ cat /home/hduser1/.ssh/id_rsa.pub >> /home/hduser1/.ssh/author
ized-keys
hduser1@ubuntu:~$ ls
Desktop Downloads hadoop-2.7.0.tar.gz Pictures Templates
```

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

A terminal window showing the execution of 'ssh localhost'. It prompts for a password and shows the warning about permanently adding 'localhost' to the list of known hosts. Then, it shows the output of 'sudo apt-get update', which lists several updates and their sizes.

```
hduser1@ubuntu:~$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:ErM+SasUz2rzrvf+3NBt6w+6w1XGyLP/Dz0dXQP70ZQ.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
hduser1@localhost's password:
Welcome to Ubuntu 16.04 LTS (GNU/Linux 4.4.0-21-generic i686)

* Documentation:  https://help.ubuntu.com/

411 packages can be updated.
145 updates are security updates.

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

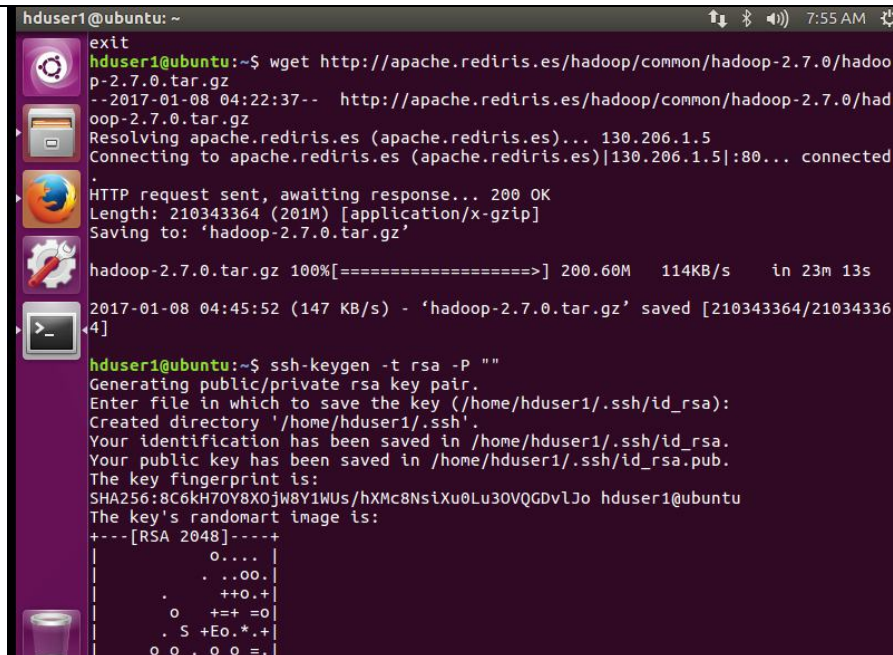
hduser1@ubuntu:~$ exit
logout
Connection to localhost closed.
hduser1@ubuntu:~$ sudo apt-get update
Get:1 http://security.ubuntu.com/ubuntu xenial-security InRelease [102 kB]
Hit:2 http://us.archive.ubuntu.com/ubuntu xenial InRelease
Get:3 http://us.archive.ubuntu.com/ubuntu xenial-updates InRelease [102 kB]
Hit:4 http://ppa.launchpad.net/webupd8team/java/ubuntu xenial InRelease
Get:5 http://us.archive.ubuntu.com/ubuntu xenial-backports InRelease [102 kB]
Fetched 306 kB in 9s (31.9 kB/s)
Reading package lists... Done
hduser1@ubuntu:~$ mv hadoop-2.7.0 hadoop
```

# Hadoop Installation

## 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
```



```
hduser1@ubuntu: ~
exit
hduser1@ubuntu:~$ wget http://apache.rediris.es/hadoop/common/hadoop-2.7.0/hadoop-2.7.0.tar.gz
--2017-01-08 04:22:37-- http://apache.rediris.es/hadoop/common/hadoop-2.7.0/hadoop-2.7.0.tar.gz
Resolving apache.rediris.es (apache.rediris.es)... 130.206.1.5
Connecting to apache.rediris.es (apache.rediris.es)|130.206.1.5|:80... connected
HTTP request sent, awaiting response... 200 OK
Length: 210343364 (201M) [application/x-gzip]
Saving to: 'hadoop-2.7.0.tar.gz'

hadoop-2.7.0.tar.gz 100%[=====] 200.60M 114KB/s in 23m 13s

2017-01-08 04:45:52 (147 KB/s) - 'hadoop-2.7.0.tar.gz' saved [210343364/210343364]

hduser1@ubuntu:~$ ssh-keygen -t rsa -P ""
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hduser1/.ssh/id_rsa):
Created directory '/home/hduser1/.ssh'.
Your identification has been saved in /home/hduser1/.ssh/id_rsa.
Your public key has been saved in /home/hduser1/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:8C6kH7OY8X0jW8Y1WUs/hXMc8NsIXu0Lu30VQGDvLJo hduser1@ubuntu
The key's randomart image is:
+---[RSA 2048]---+
|  o.... |
|  .oo. |
|  ++o.+ |
|  o  +=+ =o |
|  . S +Eo.*.+ |
|  o o . o o =. |
+---+-----+

```

## 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.

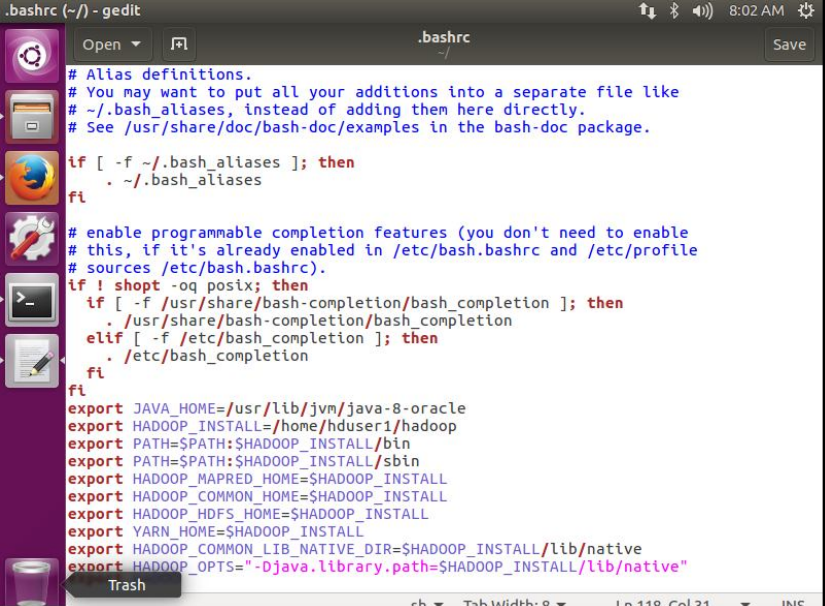
**a) Add to ~/.bashrc file:**

```
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"
```

# Hadoop Installation



```
.bashrc (~) - gedit
# Alias definitions.
# You may want to put all your additions into a separate file like
# ~/.bash_aliases, instead of adding them here directly.
# See /usr/share/doc/bash-doc/examples in the bash-doc package.

if [ -f ~/.bash_aliases ]; then
    . ~/.bash_aliases
fi

# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
    if [ -f /usr/share/bash-completion/bash_completion ]; then
        . /usr/share/bash-completion/bash_completion
    elif [ -f /etc/bash_completion ]; then
        . /etc/bash_completion
    fi
fi

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.xml to 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
```



# Hadoop Installation

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

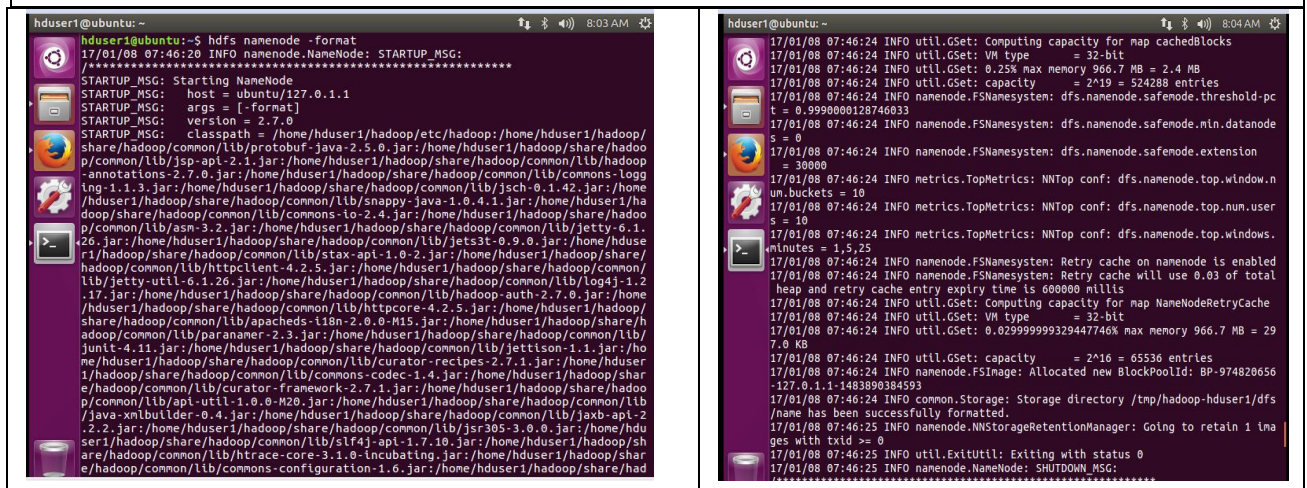
```
<configuration>
<property>
<name>mapreduce.framework.name</name>
<value>yarn</value>
</property>
</configuration>
```

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

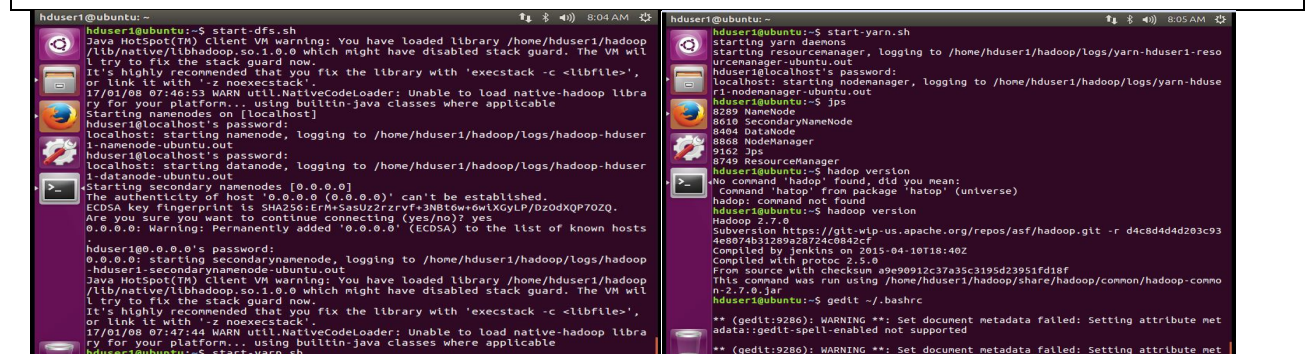


```
hduser1@ubuntu:~$ hdfs namenode -format
17/01/08 07:46:22 INFO namenode.NameNode: STARTUP_MSG:
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = ubuntu127.0.1.1
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 2.7.0
STARTUP_MSG: classpath = /home/hduser1/hadoop/etc/hadoop:/home/hduser1/hadoop/
share/hadoop/common/lib/protobuf-java-2.5.0.jar:/home/hduser1/hadoop/sha
re/hadoop/common/lib/jsp-api-2.1.jar:/home/hduser1/hadoop/share/hadoop
p/common/lib/jsp-api-2.1.jar:/home/hduser1/hadoop/share/hadoop/common/lib/hadoop
-annotations-2.7.0.jar:/home/hduser1/hadoop/share/hadoop/common/lib/commons-logg
ing-1.1.3.jar:/home/hduser1/hadoop/share/hadoop/common/lib/jsch-0.1.42.jar:/home
/hduser1/hadoop/share/hadoop/common/lib/snappy-java-1.0.4.1.jar:/home/hduser1/ha
doo/share/hadoop/common/lib/commons-io-2.4.jar:/home/hduser1/hadoop/share/hadoo
p/common/lib/asn1-3.2.jar:/home/hduser1/hadoop/share/hadoop/common/lib/jetty-6.1
.26.jar:/home/hduser1/hadoop/share/hadoop/common/lib/jets3t-0.9.0.jar:/home/hduse
r1/hadoop/share/hadoop/common/lib/stax-api-1.0-2.jar:/home/hduser1/hadoop/share/
hadoop/common/lib/httpclient-4.2.5.jar:/home/hduser1/hadoop/share/hadoop/common/
lib/jetty-utl-6.1.26.jar:/home/hduser1/hadoop/share/hadoop/common/lib/log4j-1.2
.17.jar:/home/hduser1/hadoop/share/hadoop/common/lib/hadoop-auth-2.7.0.jar:/home
/hduser1/hadoop/share/hadoop/common/lib/httpcore-4.2.5.jar:/home/hduser1/hadoop/
share/hadoop/common/lib/apacheds-i18n-2.0.0-M15.jar:/home/hduser1/hadoop/share/h
adoop/common/lib/paranamer-2.3.jar:/home/hduser1/hadoop/share/hadoop/common/lib/
junit-4.11.jar:/home/hduser1/hadoop/share/hadoop/common/lib/jettison-1.1.jar:/ho
me/hduser1/hadoop/share/hadoop/common/lib/curator-recipes-2.7.1.jar:/home/hduser
1/hadoop/share/hadoop/common/lib/commons-codec-1.4.jar:/home/hduser1/hadoop/shar
e/hadoop/common/lib/curator-framework-2.7.1.jar:/home/hduser1/hadoop/share/hadoo
p/common/lib/api-uttl-1.0.0-M20.jar:/home/hduser1/hadoop/share/hadoop/common/lib
/java-xmlbuilder-0.4.jar:/home/hduser1/hadoop/share/hadoop/common/lib/jaxb-api-2
.2.2.jar:/home/hduser1/hadoop/share/hadoop/common/lib/jsr305-3.0.0.jar:/home/hdu
ser1/hadoop/share/hadoop/common/lib/slf4j-api-1.7.10.jar:/home/hduser1/hadoop/sh
are/hadoop/common/lib/htrace-core-3.1.0-incubating.jar:/home/hduser1/hadoop/shar
e/hadoop/common/lib/commons-configuration-1.6.jar:/home/hduser1/hadoop/share/had
```

```
hduser1@ubuntu:~$ hdfs namenode -format
17/01/08 07:46:24 INFO util.GSet: Computing capacity for map cachedBlocks
17/01/08 07:46:24 INFO util.GSet: VM type = 32-bit
17/01/08 07:46:24 INFO util.GSet: 0.25% max memory 966.7 MB = 2.4 MB
17/01/08 07:46:24 INFO util.GSet: capacity = 2^19 = 524288 entries
17/01/08 07:46:24 INFO namenode.FSNamesystem: dfs.namenode.safemode.threshold-pc
t = 0.9990000128746033
17/01/08 07:46:24 INFO namenode.FSNamesystem: dfs.namenode.safemode.min.datanode
s = 0
17/01/08 07:46:24 INFO namenode.FSNamesystem: dfs.namenode.safemode.extension
= 30000
17/01/08 07:46:24 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.n
um.buckets = 10
17/01/08 07:46:24 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.user
s = 10
17/01/08 07:46:24 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.
minutes = 15,25
17/01/08 07:46:24 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
17/01/08 07:46:24 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total
heap and retry cache entry expiry time is 600000 millis
17/01/08 07:46:24 INFO util.GSet: Computing capacity for map NameNodeRetryCache
17/01/08 07:46:24 INFO util.GSet: VM type = 32-bit
17/01/08 07:46:24 INFO util.GSet: 0.029999999329447746% max memory 966.7 MB = 29
7.8 MB
17/01/08 07:46:24 INFO util.GSet: capacity = 2^16 = 65536 entries
17/01/08 07:46:24 INFO namenode.FSInage: Allocated new BlockPoolId: BP-974820656-
127.0.1.1-1483890384593
17/01/08 07:46:24 INFO common.Storage: Storage directory /tmp/hadoop-hduser1/dfs
/name has been successfully formatted.
17/01/08 07:46:25 INFO namenode.NNStorageRetentionManager: Going to retain 1 ina
ges with txid >= 0
17/01/08 07:46:25 INFO util.ExitUtil: Exiting with status 0
17/01/08 07:46:25 INFO namenode.NameNode: SHUTDOWN_MSG:
*****
```

## Start Hadoop:

### start-all.sh



```
hduser1@ubuntu:~$ start-all.sh
Java HotSpot(TM) client VM warning: You have loaded library /home/hduser1/hadoop
/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will
try to fix the stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>',
or link it with '-z noexecstack'.
17/01/08 07:46:53 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
hduser1@localhost:~$ password:
localhost: starting namenode, logging to /home/hduser1/hadoop/logs/hadoop-hduse
r1-namenode-ubuntu.out
hduser1@localhost:~$ password:
localhost: starting datanode, logging to /home/hduser1/hadoop/logs/hadoop-hduse
r1-datanode-ubuntu.out
Starting secondary namenodes [0.0.0.0]
The authenticity of host '0.0.0.0 (0.0.0.0)' can't be established.
ECDSA key fingerprint is SHA256:ERm+SasU22rzrVf3N8t6w+6WtXGyLP/D20dXQP7020.
Are you sure you want to continue connecting (yes/no)? yes
0.0.0.0: Warning: Permanently added '0.0.0.0' (ECDSA) to the list of known hosts
hduser1@0.0.0.0's password:
0.0.0.0: starting secondarynamenode, logging to /home/hduser1/hadoop/logs/hadoop
-hduser1-secondarynamenode-ubuntu.out
Java HotSpot(TM) client VM warning: You have loaded library /home/hduser1/hadoop
/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will
try to fix the stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>',
or link it with '-z noexecstack'.
17/01/08 07:47:44 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
hduser1@ubuntu:~$ start-yarn.sh
hduser1@ubuntu:~$ start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /home/hduser1/hadoop/logs/yarn-hduser1-reso
urcemanag
hduser1@localhost:~$ password:
localhost: starting nodemanager, logging to /home/hduser1/hadoop/logs/yarn-hduse
r1-nodemanager-ubuntu.out
hduser1@localhost:~$ password:
localhost: starting nodemanager, logging to /home/hduser1/hadoop/logs/yarn-hduse
r1-nodemanager-ubuntu.out
8289 NameNode
8610 SecondaryNameNode
8404 DataNode
8868 NodeManager
9162 Jps
8749 ResourceManager
hduser1@ubuntu:~$ hadoop version
No command 'hadoop' found, did you mean:
Command 'hadoop' from package 'hadoop' (universe)
Compiled by Jenkins on 2015-04-10T18:40Z
hduser1@ubuntu:~$ hadoop version
Hadoop 2.7.0
Subversion https://git-wip-us.apache.org/repos/asf/hadoop.git -r d4c8d4d4d203c93
4e807b31289a28724c0842cf
Compiled by Jenkins on 2015-04-10T18:40Z
From source with checksum a9e90912c37a35c3195d23951fd18f
This command was run using /home/hduser1/hadoop/share/hadoop/common/hadoop-commo
n-2.7.0.jar
hduser1@ubuntu:~$ gedit ~/.bashrc
** (gedit:9286): WARNING **: Set document metadata failed: Setting attribute met
adata: gedit:spell-enabled not supported
** (gedit:9286): WARNING **: Set document metadata failed: Setting attribute met
```

# Hadoop Installation

**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>

