

Single Node Apache Hadoop Setup and Configuration on Ubuntu 14.04



Data Science Lab, The Department of Computer Science,
KSKV Kachchh University.

Web: <http://cs.kutchuni.edu.in>

The MIT License (MIT)

Copyright (c) 2016. Data Science Lab, University of Kachchh.

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

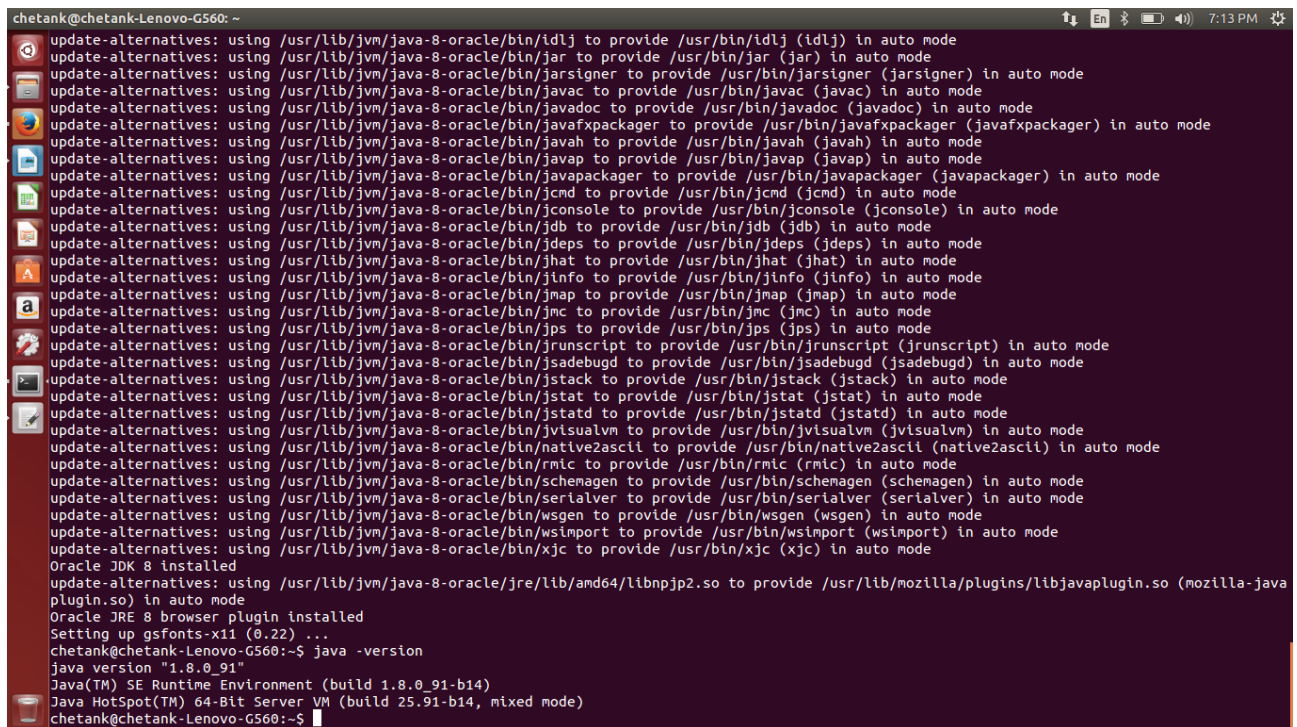
- 1) Ubuntu 14.04 64bit Installed.
- 2) Java 1.8 Installed with following steps

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

To automatically set up the Java 8 environment variables

Check it

```
java -version
```



```
chetan@chetan-Lenovo-G560: ~
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/idlj to provide /usr/bin/idlj (idlj) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jar to provide /usr/bin/jar (jar) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jarsigner to provide /usr/bin/jarsigner (jarsigner) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/javac to provide /usr/bin/javac (javac) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/javadoc to provide /usr/bin/javadoc (javadoc) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/javafxpackager to provide /usr/bin/javafxpackager (javafxpackager) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/javah to provide /usr/bin/javah (javah) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/javap to provide /usr/bin/javap (javap) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/javapackager to provide /usr/bin/javapackager (javapackager) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jcmd to provide /usr/bin/jcmd (jcmd) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jconsole to provide /usr/bin/jconsole (jconsole) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jdb to provide /usr/bin/jdb (jdb) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jdeps to provide /usr/bin/jdeps (jdeps) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jhat to provide /usr/bin/jhat (jhat) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jinfo to provide /usr/bin/jinfo (jinfo) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jmap to provide /usr/bin/jmap (jmap) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jmc to provide /usr/bin/jmc (jmc) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jps to provide /usr/bin/jps (jps) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jrunscript to provide /usr/bin/jrunscript (jrunscript) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jsadebugd to provide /usr/bin/jsadebugd (jsadebugd) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jstack to provide /usr/bin/jstack (jstack) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jstat to provide /usr/bin/jstat (jstat) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jstatd to provide /usr/bin/jstatd (jstatd) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/jvisualvm to provide /usr/bin/jvisualvm (jvisualvm) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/native2ascii to provide /usr/bin/native2ascii (native2ascii) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/rmic to provide /usr/bin/rmic (rmic) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/schemagen to provide /usr/bin/schemagen (schemagen) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/serialver to provide /usr/bin/serialver (serialver) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/wsgen to provide /usr/bin/wsgen (wsgen) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/wsimport to provide /usr/bin/wsimport (wsimport) in auto mode
update-alternatives: using /usr/lib/jvm/java-8-oracle/bin/xjc to provide /usr/bin/xjc (xjc) in auto mode
Oracle JDK 8 installed
update-alternatives: using /usr/lib/jvm/java-8-oracle/jre/lib/amd64/libnptl2.so to provide /usr/lib/mono/plugins/libjavaplugin.so (mozilla-java
plugin.so) in auto mode
Oracle JRE 8 browser plugin installed
Setting up gsfonnts-x11 (0.22) ...
chetan@chetan-Lenovo-G560:~$ java -version
java version "1.8.0_91"
Java(TM) SE Runtime Environment (build 1.8.0_91-b14)
Java HotSpot(TM) 64-Bit Server VM (build 25.91-b14, mixed mode)
chetan@chetan-Lenovo-G560:~$
```

[Figure 1] Installation of JAVA 1.8 on Ubuntu 14.04

Set JAVA_HOME System environment variable, follow following steps.

1. Sudo gedit ~/.bashrc
2. go to end of the file, write *export JAVA_HOME=/usr/lib/jvm/java-8-oracle*
3. save & close the bashrc file
4. source ~/.bashrc (for saving environment variable permanently)
5. for check – echo \$JAVA_HOME will give you entire path.

```
chettank@chettank-Lenovo-G560: /usr/lib/jvm/java-8-oracle
chettank@chettank-Lenovo-G560:/usr/lib/jvm/java-8-oracle$ sudo gedit ~/.bashrc
(gedit:4820): Gtk-WARNING **: Calling Inhibit failed: GDBus.Error:org.freedesktop.DBus.Error.ServiceUnknown: The name org.gnome.SessionManager was not provided by any .service files
(gedit:4820): Gtk-WARNING **: Calling Inhibit failed: GDBus.Error:org.freedesktop.DBus.Error.ServiceUnknown: The name org.gnome.SessionManager was not provided by any .service files
^Cchettank@chettank-Lenovo-G560:/usr/lib/jvm/java-8-oracle$ source ~/.bashrc
chettank@chettank-Lenovo-G560:/usr/lib/jvm/java-8-oracle$ echo $JAVA_HOME
/usr/lib/jvm/java-8-oracle
chettank@chettank-Lenovo-G560:/usr/lib/jvm/java-8-oracle$
```

[Figure 2] Checking JAVA_HOME environment variable

3) Installation of Hadoop 2.7.0

Download Hadoop 2.7.0

wget <https://archive.apache.org/dist/hadoop/core/hadoop-2.7.0/hadoop-2.7.0.tar.gz>

```
chetak@chetak-Lenovo-G560: ~/Downloads
google-chrome-stable_current_amd64.deb  spark-1.6.1-bin-without-hadoop.tgz
hadoop-2.7.0-src/                        spark-1.6.1-bin-without-hadoop.tgz.part
hadoop-2.7.0-src.tar.gz
chetak@chetak-Lenovo-G560:~/Downloads$ rm -rf hadoop-2.7.0-src
chetak@chetak-Lenovo-G560:~/Downloads$ rm -rf hadoop-2.7.0-src.tar.gz
chetak@chetak-Lenovo-G560:~/Downloads$ l
google-chrome-stable_current_amd64.deb  spark-1.6.1-bin-without-hadoop.tgz
chetak@chetak-Lenovo-G560:~/Downloads$ gwet https://archive.apache.org/dist/h
doop/core/hadoop-2.7.0/hadoop-2.7.0.tar.gz
No command 'gwet' found, did you mean:
  Command 'wget' from package 'wget' (main)
gwet: command not found
chetak@chetak-Lenovo-G560:~/Downloads$ wget https://archive.apache.org/dist/h
doop/core/hadoop-2.7.0/hadoop-2.7.0.tar.gz
--2016-06-30 19:30:58--  https://archive.apache.org/dist/hadoop/core/hadoop-2.7
0/hadoop-2.7.0.tar.gz
Resolving archive.apache.org (archive.apache.org)... 163.172.17.199
Connecting to archive.apache.org (archive.apache.org)|163.172.17.199|:443... co
nected.
HTTP request sent, awaiting response... 200 OK
Length: 210343364 (201M) [application/x-gzip]
Saving to: 'hadoop-2.7.0.tar.gz'

33% [=====>                               ] 6,99,02,336 130KB/s  eta 2m 54s
```

[Figure 3] Downloading Apache Hadoop 2.7.0

Uncompress the tar file and move to /usr/local/hadoop

```
$ tar xzf hadoop-2.7.0.tar.gz
```

```
$ sudo mv hadoop-2.7.0 /usr/local/hadoop
```

4) Add Hadoop Group and User

To avoid security issues, it's always a good practice to setup new Hadoop user group and user account to deal with all Hadoop related activities. We will create **hadoop** as system group and **hduser** as system user.

```
$ sudo addgroup hadoop
```

```
$ sudo adduser --ingroup hadoop hduser
```

```
$ sudo adduser hduser sudo
```

```
cheta...@cheta...-Lenovo-G560: ~/Downloads
hadoop-2.7.0/LICENSE.txt
cheta...@cheta...-Lenovo-G560:~/Downloads$ l
eclipse-jee-indigo-SR2-linux-gtk-x86_64.tar.gz
eclipse-jee-indigo-SR2-linux-gtk-x86_64.tar.gz.part
google-chrome-stable_current_amd64.deb
hadoop-2.7.0/
hadoop-2.7.0.tar.gz
spark-1.6.1-bin-without-hadoop.tgz
cheta...@cheta...-Lenovo-G560:~/Downloads$ mv hadoop-2.7.0 /usr/local/hadoop
mv: cannot move 'hadoop-2.7.0' to '/usr/local/hadoop': Permission denied
cheta...@cheta...-Lenovo-G560:~/Downloads$ chmod 644 /usr/local/hadoop
chmod: cannot access '/usr/local/hadoop': No such file or directory
cheta...@cheta...-Lenovo-G560:~/Downloads$ sudo chmod 644 /usr/local/hadoop
[sudo] password for cheta...:
cheta...@cheta...-Lenovo-G560:~/Downloads$ sudo mv hadoop-2.7.0 /usr/local/hadoop
[sudo] password for cheta...:
cheta...@cheta...-Lenovo-G560:~/Downloads$ sudo addgroup hadoop
Adding group 'hadoop' (GID 1001) ...
Done.
cheta...@cheta...-Lenovo-G560:~/Downloads$ sudo adduser --ingroup hadoop hduser
Adding user 'hduser' ...
Adding new user 'hduser' (1001) with group 'hadoop' ...
Creating home directory '/home/hduser' ...
Copying files from '/etc/skel' ...
Enter new UNIX password:
Retype new UNIX password:
passwd: password updated successfully
Changing the user information for hduser
Enter the new value, or press ENTER for the default
  Full Name []:
  Room Number []:
  Work Phone []:
  Home Phone []:
  Other []:
Is the information correct? [Y/n] Y
cheta...@cheta...-Lenovo-G560:~/Downloads$ sudo adduser hduser sudo
Adding user 'hduser' to group 'sudo' ...
Adding user hduser to group sudo
Done.
cheta...@cheta...-Lenovo-G560:~/Downloads$
```

[Figure 4] Adding Hadoop group and hduser in ubuntu

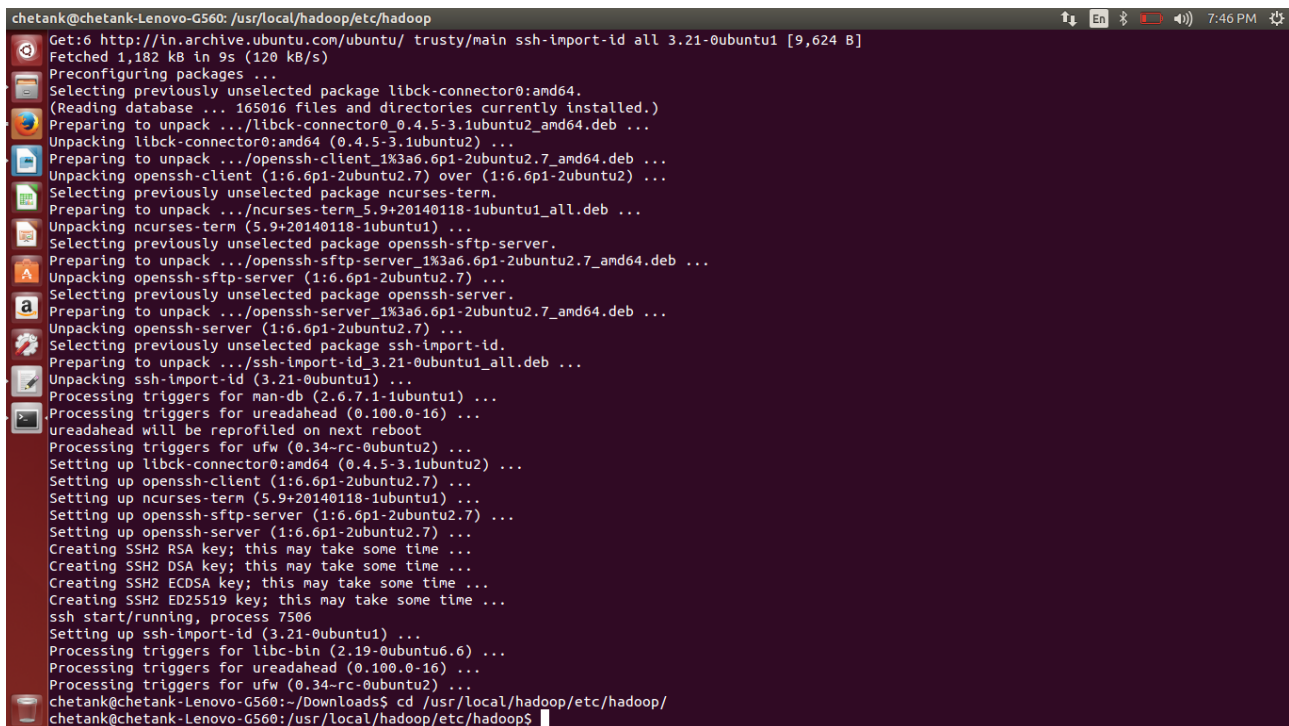
5) Installing SSH

SSH (“Secure SHell”) is a protocol for securely accessing one machine from another. Hadoop uses SSH for accessing another slaves nodes to start and manage all HDFS and MapReduce daemons. (i.e Daemons – Build it processes of the application service)

\$sudo apt-get install openssh-server

```
cheta...@cheta...-Lenovo-G560: ~/Downloads
cheta...@cheta...-Lenovo-G560:~/Downloads$ l
eclipse-jee-indigo-SR2-linux-gtk-x86_64.tar.gz
eclipse-jee-indigo-SR2-linux-gtk-x86_64.tar.gz.part
google-chrome-stable_current_amd64.deb
hadoop-2.7.0/
hadoop-2.7.0.tar.gz
spark-1.6.1-bin-without-hadoop.tgz
cheta...@cheta...-Lenovo-G560:~/Downloads$ mv hadoop-2.7.0 /usr/local/hadoop
mv: cannot move 'hadoop-2.7.0' to '/usr/local/hadoop': Permission denied
cheta...@cheta...-Lenovo-G560:~/Downloads$ chmod 644 /usr/local/hadoop
chmod: cannot access '/usr/local/hadoop': No such file or directory
cheta...@cheta...-Lenovo-G560:~/Downloads$ sudo chmod 644 /usr/local/hadoop
[sudo] password for cheta...:
cheta...@cheta...-Lenovo-G560:~/Downloads$ sudo mv hadoop-2.7.0 /usr/local/hadoop
[sudo] password for cheta...:
cheta...@cheta...-Lenovo-G560:~/Downloads$ sudo addgroup hadoop
Adding group 'hadoop' (GID 1001) ...
Done.
cheta...@cheta...-Lenovo-G560:~/Downloads$ sudo adduser --ingroup hadoop hduser
Adding user 'hduser' ...
Adding new user 'hduser' (1001) with group 'hadoop' ...
Creating home directory '/home/hduser' ...
Copying files from '/etc/skel' ...
Enter new UNIX password:
Retype new UNIX password:
passwd: password updated successfully
Changing the user information for hduser
Enter the new value, or press ENTER for the default
  Full Name []:
  Room Number []:
  Work Phone []:
  Home Phone []:
  Other []:
Is the information correct? [Y/n] Y
cheta...@cheta...-Lenovo-G560:~/Downloads$ sudo adduser hduser sudo
Adding user 'hduser' to group 'sudo' ...
Adding user hduser to group sudo
Done.
cheta...@cheta...-Lenovo-G560:~/Downloads$ sudo apt-get install openssh-server
Reading package lists... 98%
```

[Figure 5] Install openssh-server



```
chetak@chetak-Lenovo-G560: /usr/local/hadoop/etc/hadoop
Get:6 http://in.archive.ubuntu.com/ubuntu/ trusty/main ssh-import-id all 3.21-0ubuntu1 [9,624 B]
Fetched 1,182 kB in 9s (120 kB/s)
Preconfiguring packages ...
Selecting previously unselected package libck-connector0:amd64.
(Reading database ... 165016 files and directories currently installed.)
Preparing to unpack .../libck-connector0_0.4.5-3.1ubuntu2_amd64.deb ...
Unpacking libck-connector0:amd64 (0.4.5-3.1ubuntu2) ...
Preparing to unpack .../openssh-client_1:6.6p1-2ubuntu2.7_amd64.deb ...
Unpacking openssh-client (1:6.6p1-2ubuntu2.7) over (1:6.6p1-2ubuntu2) ...
Selecting previously unselected package ncurses-term.
Preparing to unpack .../ncurses-term_5.9+20140118-1ubuntu1_all.deb ...
Unpacking ncurses-term (5.9+20140118-1ubuntu1) ...
Selecting previously unselected package openssh-sftp-server.
Preparing to unpack .../openssh-sftp-server_1:6.6p1-2ubuntu2.7_amd64.deb ...
Unpacking openssh-sftp-server (1:6.6p1-2ubuntu2.7) ...
Selecting previously unselected package openssh-server.
Preparing to unpack .../openssh-server_1:6.6p1-2ubuntu2.7_amd64.deb ...
Unpacking openssh-server (1:6.6p1-2ubuntu2.7) ...
Selecting previously unselected package ssh-import-id.
Preparing to unpack .../ssh-import-id_3.21-0ubuntu1_all.deb ...
Unpacking ssh-import-id (3.21-0ubuntu1) ...
Processing triggers for man-db (2.6.7.1-1ubuntu1) ...
Processing triggers for ureadahead (0.100.0-16) ...
ureadahead will be reprofiled on next reboot
Processing triggers for ufw (0.34-rc-0ubuntu2) ...
Setting up libck-connector0:amd64 (0.4.5-3.1ubuntu2) ...
Setting up openssh-client (1:6.6p1-2ubuntu2.7) ...
Setting up ncurses-term (5.9+20140118-1ubuntu1) ...
Setting up openssh-sftp-server (1:6.6p1-2ubuntu2.7) ...
Setting up openssh-server (1:6.6p1-2ubuntu2.7) ...
Creating SSH2 RSA key; this may take some time ...
Creating SSH2 DSA key; this may take some time ...
Creating SSH2 ECDSA key; this may take some time ...
Creating SSH2 ED25519 key; this may take some time ...
ssh start/running, process 7506
Setting up ssh-import-id (3.21-0ubuntu1) ...
Processing triggers for libc-bin (2.19-0ubuntu6.6) ...
Processing triggers for ureadahead (0.100.0-16) ...
Processing triggers for ufw (0.34-rc-0ubuntu2) ...
chetak@chetak-Lenovo-G560:~/Downloads$ cd /usr/local/hadoop/etc/hadoop/
chetak@chetak-Lenovo-G560: /usr/local/hadoop/etc/hadoop$
```

[Figure 6] Install openssh-server

6) Switch to Hadoop User account

\$ su – hduser

7) Configure Hadoop

To complete the setup of Hadoop, the following files will have to be modified:

- 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/yarn-site.xml
- 5) /usr/local/hadoop/etc/hadoop/mapred-site.xml
- 6) /usr/local/hadoop/etc/hadoop/hdfs-site.xml

Let's start configuring one by one.

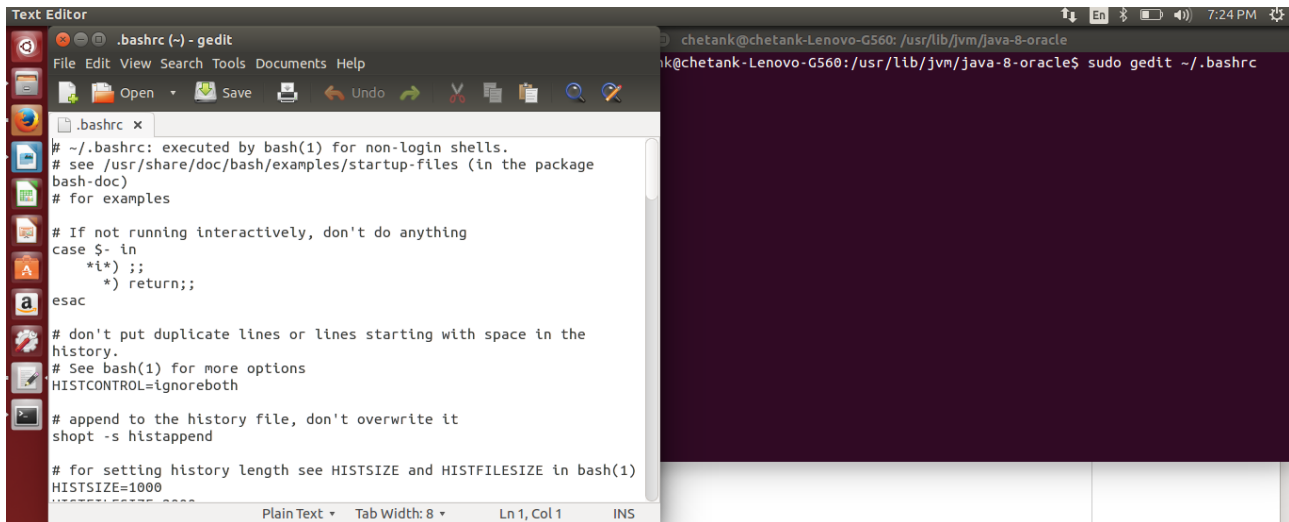
7.1) ~/.bashrc

open the ~/.bashrc and paste the following at the end. Putting the above content in the .bashrc file ensures that these variables are always available when your VPS starts up.

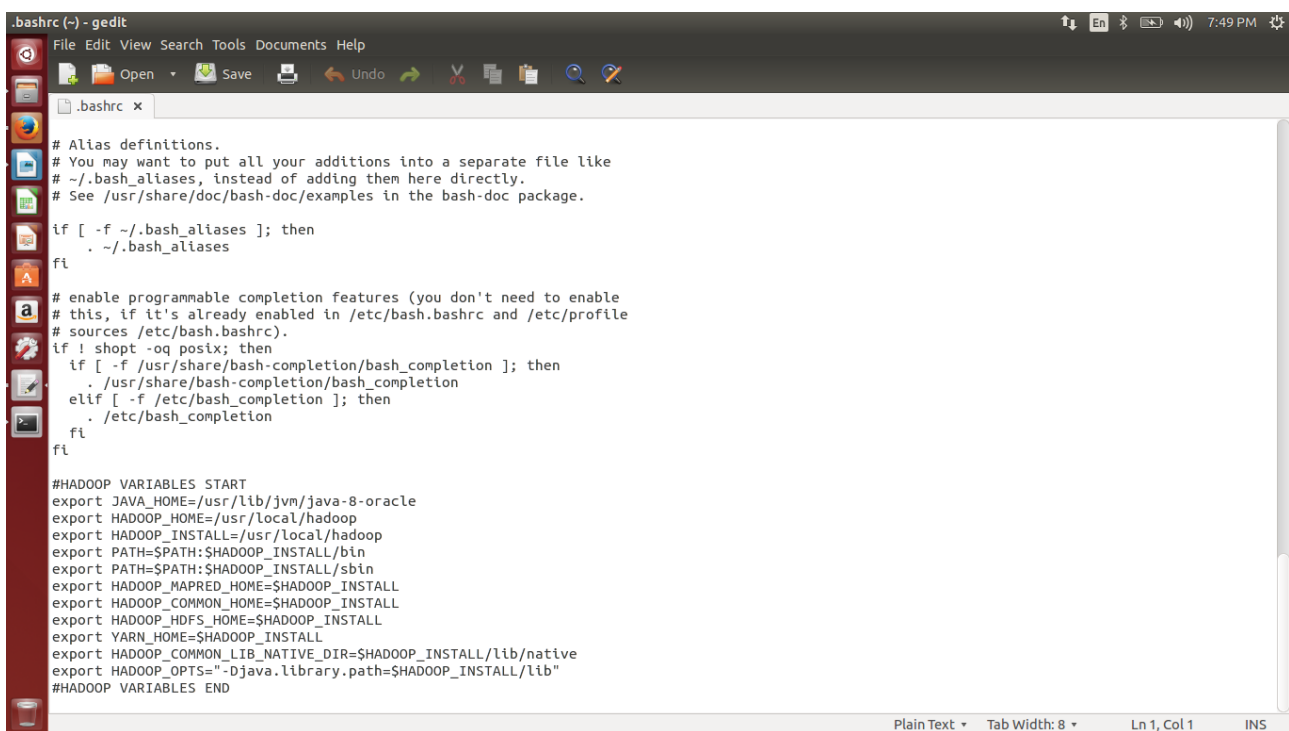
```
#HADOOP VARIABLES START
export JAVA_HOME=/usr/lib/jvm/java-8-oracle
export HADOOP_HOME=/usr/local/hadoop
export HADOOP_INSTALL=/usr/local/hadoop
export PATH=$PATH:$HADOOP_INSTALL/bin
export PATH=$PATH:$HADOOP_INSTALL/sbin
export HADOOP_MAPRED_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_HOME=$HADOOP_INSTALL
```



```
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
```



[Figure 7] Configure environment variable with ~/.bashrc



[Figure 8] Hadoop environment variables with ~/.bashrc

A terminal window titled 'chetank@chetank-Lenovo-G560: /usr/lib/jvm/java-8-oracle'. The user enters 'sudo gedit ~/.bashrc', opening a gedit editor window. The terminal shows two GTK warnings from gedit: '(gedit:4820): Gtk-WARNING **: Calling Inhibit failed: GDBus.Error:org.freedesktop.DBus.Error.ServiceUnknown: The name org.gnome.SessionManager was not provided by any .service files'. The user then enters '^C' to exit gedit, followed by 'source ~/.bashrc' to reload the environment variables. The prompt returns to 'chetank@chetank-Lenovo-G560: /usr/lib/jvm/java-8-oracle\$' with a cursor.

```
chetank@chetank-Lenovo-G560: /usr/lib/jvm/java-8-oracle$ sudo gedit ~/.bashrc
(gedit:4820): Gtk-WARNING **: Calling Inhibit failed: GDBus.Error:org.freedesktop.DBus.Error.ServiceUnknown: The name org.gnome.SessionManager was not provided by any .service files
(gedit:4820): Gtk-WARNING **: Calling Inhibit failed: GDBus.Error:org.freedesktop.DBus.Error.ServiceUnknown: The name org.gnome.SessionManager was not provided by any .service files
^Cchetank@chetank-Lenovo-G560: /usr/lib/jvm/java-8-oracle$ source ~/.bashrc
chetank@chetank-Lenovo-G560: /usr/lib/jvm/java-8-oracle$
```

[Figure 9] Save environment variables with ~/.bashrc

7.2) Edit `hadoop-env.sh`

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

Change `JAVA_HOME` variable into

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

7.3) `core-site.xml`

`/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. Enter the following content in between the tag `<configuration></configuration>`

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



```
cheta...@cheta...-Lenovo-G560: /usr/local/hadoop/etc/hadoop$ ls
capacity-scheduler.xml  hadoop-metrics2.properties  httpfs-signature.secret  log4j.properties          ssl-client.xml.example
configuration.xml       hadoop-metrics.properties   httpfs-site.xml          mapred-env.cmd            ssl-server.xml.example
container-executor.cfg  hadoop-policy.xml           kms-acls.xml             mapred-env.sh             yarn-env.cmd
core-site.xml           hdfs-site.xml              kms-env.sh               mapred-queues.xml.template yarn-env.sh
hadoop-env.cmd          httpfs-env.sh              kms-log4j.properties     mapred-site.xml.template  yarn-site.xml
hadoop-env.sh           httpfs-log4j.properties     kms-site.xml             slaves
cheta...@cheta...-Lenovo-G560: /usr/local/hadoop/etc/hadoop$ ls -ll
total 152
-rw-r--r-- 1 cheta... cheta... 4436 Apr 11 2015 capacity-scheduler.xml
-rw-r--r-- 1 cheta... cheta... 1335 Apr 11 2015 configuration.xml
-rw-r--r-- 1 cheta... cheta... 318 Apr 11 2015 container-executor.cfg
-rw-r--r-- 1 cheta... cheta... 774 Apr 11 2015 core-site.xml
-rw-r--r-- 1 cheta... cheta... 3670 Apr 11 2015 hadoop-env.cmd
-rw-r--r-- 1 cheta... cheta... 4224 Apr 11 2015 hadoop-env.sh
-rw-r--r-- 1 cheta... cheta... 2598 Apr 11 2015 hadoop-metrics2.properties
-rw-r--r-- 1 cheta... cheta... 2490 Apr 11 2015 hadoop-metrics.properties
-rw-r--r-- 1 cheta... cheta... 9683 Apr 11 2015 hadoop-policy.xml
-rw-r--r-- 1 cheta... cheta... 775 Apr 11 2015 hdfs-site.xml
-rw-r--r-- 1 cheta... cheta... 1449 Apr 11 2015 httpfs-env.sh
-rw-r--r-- 1 cheta... cheta... 1657 Apr 11 2015 httpfs-log4j.properties
-rw-r--r-- 1 cheta... cheta... 21 Apr 11 2015 httpfs-signature.secret
-rw-r--r-- 1 cheta... cheta... 620 Apr 11 2015 httpfs-site.xml
-rw-r--r-- 1 cheta... cheta... 3518 Apr 11 2015 kms-acls.xml
-rw-r--r-- 1 cheta... cheta... 1527 Apr 11 2015 kms-env.sh
-rw-r--r-- 1 cheta... cheta... 1631 Apr 11 2015 kms-log4j.properties
-rw-r--r-- 1 cheta... cheta... 5511 Apr 11 2015 kms-site.xml
-rw-r--r-- 1 cheta... cheta... 11237 Apr 11 2015 log4j.properties
-rw-r--r-- 1 cheta... cheta... 951 Apr 11 2015 mapred-env.cmd
-rw-r--r-- 1 cheta... cheta... 1383 Apr 11 2015 mapred-env.sh
-rw-r--r-- 1 cheta... cheta... 4113 Apr 11 2015 mapred-queues.xml.template
-rw-r--r-- 1 cheta... cheta... 758 Apr 11 2015 mapred-site.xml.template
-rw-r--r-- 1 cheta... cheta... 10 Apr 11 2015 slaves
-rw-r--r-- 1 cheta... cheta... 2316 Apr 11 2015 ssl-client.xml.example
-rw-r--r-- 1 cheta... cheta... 2268 Apr 11 2015 ssl-server.xml.example
-rw-r--r-- 1 cheta... cheta... 2250 Apr 11 2015 yarn-env.cmd
-rw-r--r-- 1 cheta... cheta... 4567 Apr 11 2015 yarn-env.sh
-rw-r--r-- 1 cheta... cheta... 690 Apr 11 2015 yarn-site.xml
cheta...@cheta...-Lenovo-G560: /usr/local/hadoop/etc/hadoop$ sudo gedit core-site.xml
```

[Figure 10] Opening core-site.xml from /usr/local/hadoop/etc/hadoop

```
Text Editor
hduser@cheta...-Lenovo-G560: /usr/local/hadoop/etc/hadoop$ ls
capacity-scheduler.xml  httpfs-env.sh          mapred-env.sh
configuration.xml       httpfs-log4j.properties mapred-queues.xml.template
container-executor.cfg  httpfs-signature.secret mapred-site.xml
core-site.xml           httpfs-site.xml        mapred-site.xml.template
hadoop-env.cmd          kms-acls.xml           slaves
hadoop-env.sh          kms-env.sh             ssl-client.xml.example
hadoop-metrics2.properties kms-log4j.properties  ssl-server.xml.example
hadoop-metrics.properties kms-site.xml           yarn-env.cmd
hadoop-policy.xml       log4j.properties      yarn-env.sh
hdfs-site.xml          mapred-env.cmd         yarn-site.xml
hduser@cheta...-Lenovo-G560: /usr/local/hadoop/etc/hadoop$ sudo gedit core-site.xml
[sudo] password for hduser:

core-site.xml (/usr/local/hadoop/etc/hadoop) - gedit
File Edit View Search Tools Documents Help
core-site.xml x
http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or
implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

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

[Figure 11] Content of core-site.xml after adding property in configuration.

7.4) yarn-site.xml

/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. This file can be used to override the default settings that MapReduce starts with. Enter the following content in between the tag

<configuration></configuration>

<property>

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

 <value>mapreduce_shuffle</value>

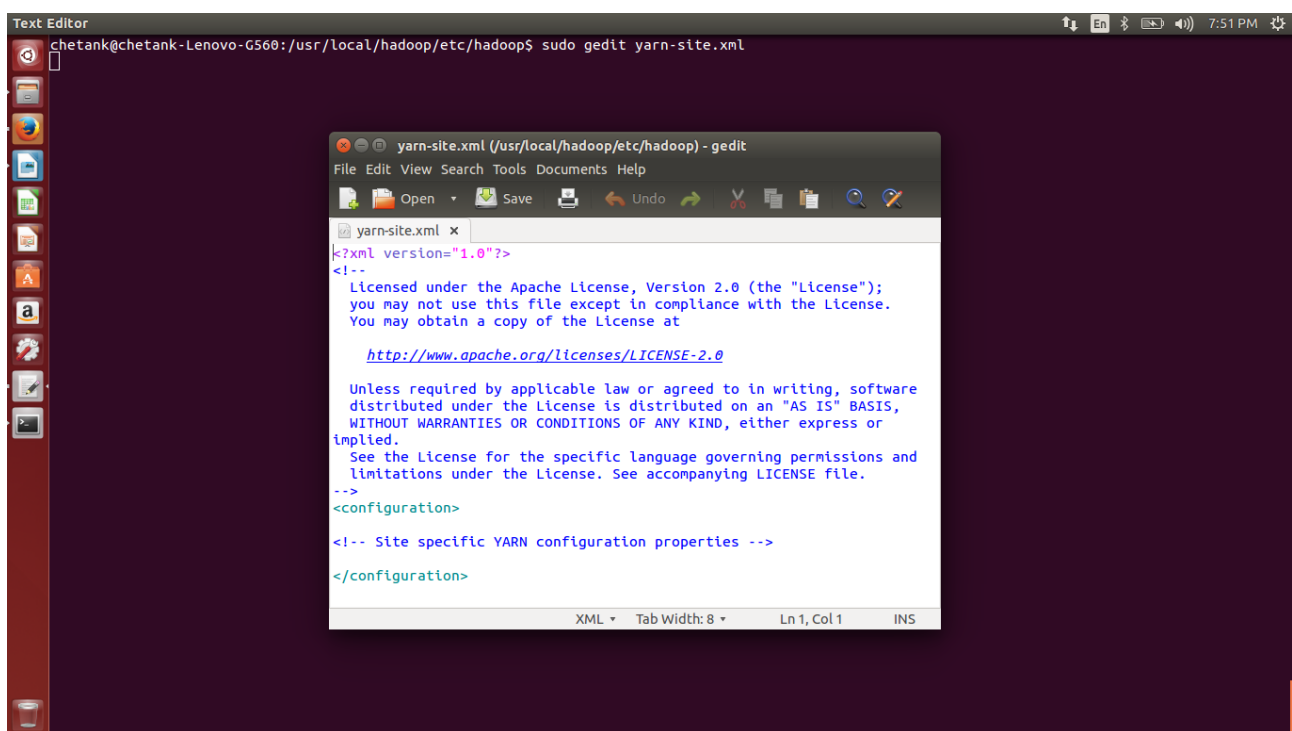
</property>

<property>

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

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

</property>



[Figure 12] Opening yarn-site.xml from /usr/local/hadoop/etc/hadoop

7.5) mapred-site.xml

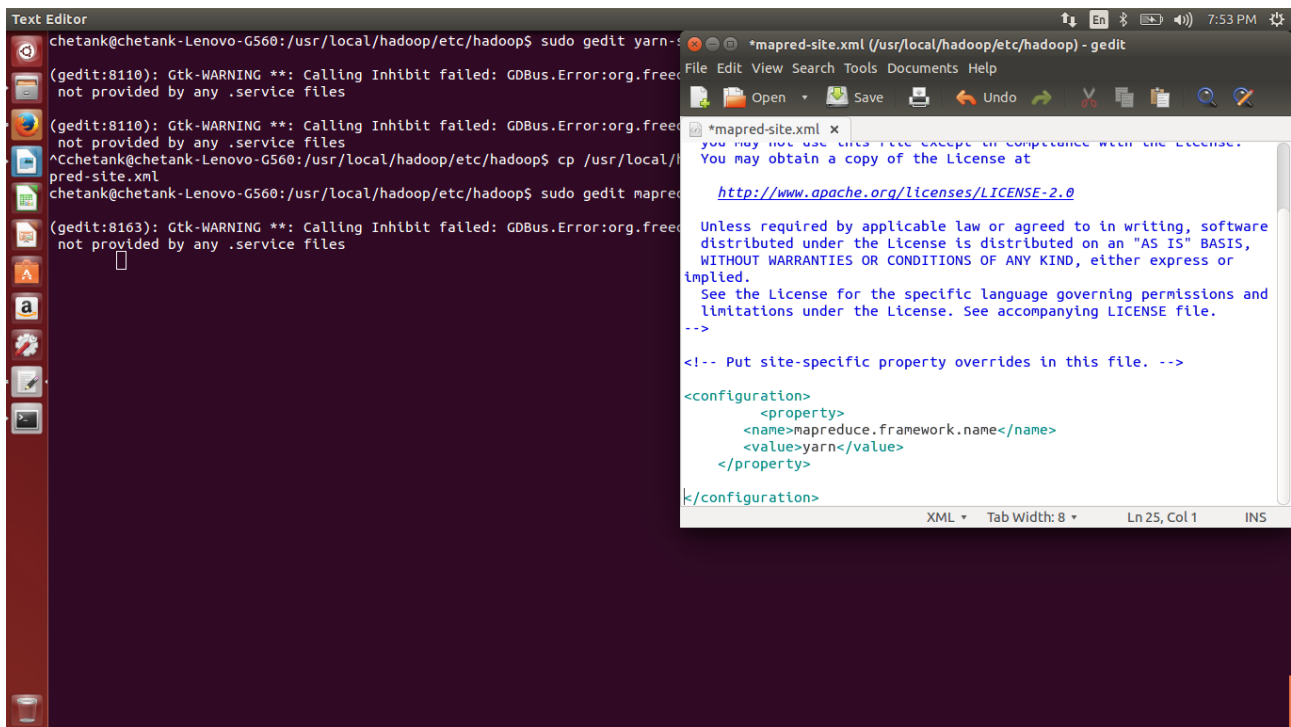
/usr/local/hadoop/etc/hadoop/mapred-site.xml

By default, the /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. This file is used to specify which framework is being used for MapReduce.

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

Then open `/usr/local/hadoop/etc/hadoop/mapred-site.xml` and enter the following content in between the tag `<configuration></configuration>`

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



[Figure 13] Configuration of mapred-site.xml

7.6) hdfs-site.xml

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

The `/usr/local/hadoop/etc/hadoop/hdfs-site.xml` has 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, we need to create two directories which will contain the namenode and the datanode for this Hadoop installation.

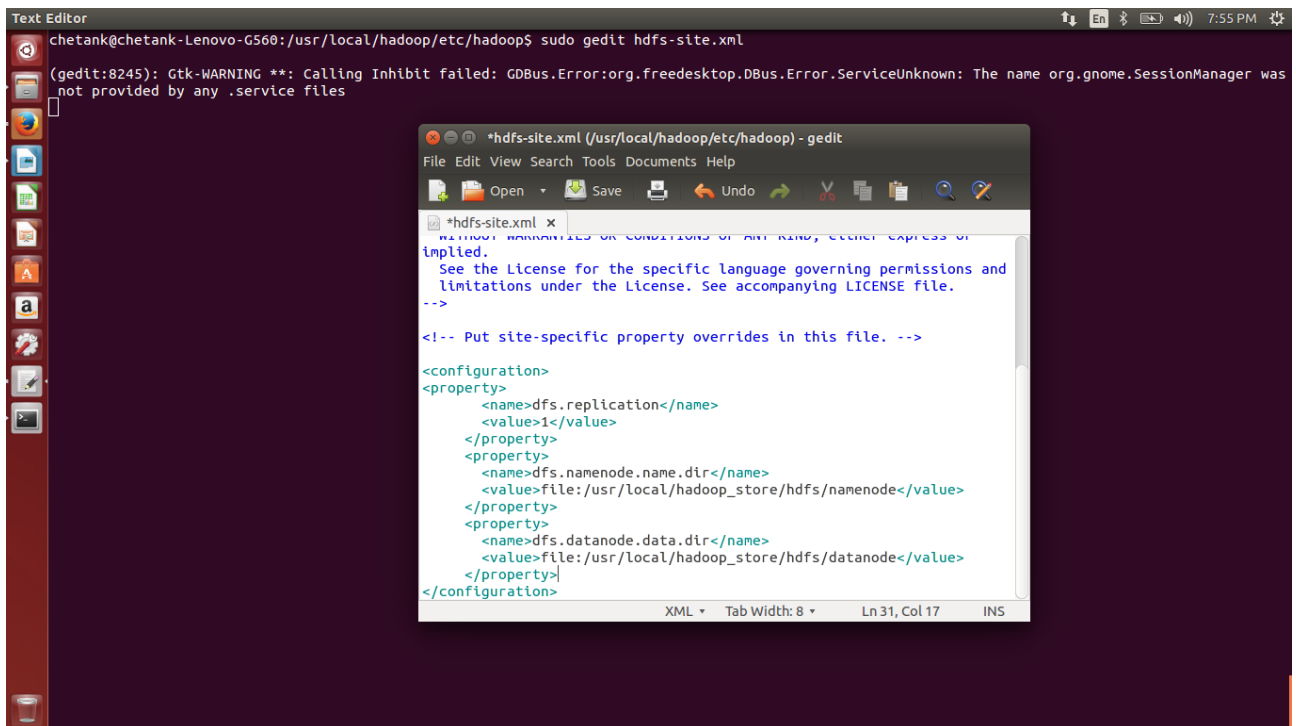
```
sudo mkdir -p /usr/local/hadoop_store/hdfs/namenode
sudo mkdir -p /usr/local/hadoop_store/hdfs/datanode
```

```
chhetank@chhetank-Lenovo-G560: /usr/local/hadoop/etc/hadoop
chhetank@chhetank-Lenovo-G560: /usr/local/hadoop/etc/hadoop$ sudo gedit yarn-site.xml
(gedit:8110): Gtk-WARNING **: Calling Inhibit failed: GDBus.Error:org.freedesktop.DBus.Error.ServiceUnknown: The name org.gnome.SessionManager was
not provided by any .service files
(gedit:8110): Gtk-WARNING **: Calling Inhibit failed: GDBus.Error:org.freedesktop.DBus.Error.ServiceUnknown: The name org.gnome.SessionManager was
not provided by any .service files
^Cchhetank@chhetank-Lenovo-G560: /usr/local/hadoop/etc/hadoop$ cp /usr/local/hadoop/etc/hadoop/napred-site.xml.template /usr/local/hadoop/etc/hadoop/
pred-site.xml
chhetank@chhetank-Lenovo-G560: /usr/local/hadoop/etc/hadoop$ sudo gedit mapred-site.xml
(gedit:8163): Gtk-WARNING **: Calling Inhibit failed: GDBus.Error:org.freedesktop.DBus.Error.ServiceUnknown: The name org.gnome.SessionManager was
not provided by any .service files
(gedit:8163): Gtk-WARNING **: Calling Inhibit failed: GDBus.Error:org.freedesktop.DBus.Error.ServiceUnknown: The name org.gnome.SessionManager was
not provided by any .service files
^Cchhetank@chhetank-Lenovo-G560: /usr/local/hadoop/etc/hadoop$
chhetank@chhetank-Lenovo-G560: /usr/local/hadoop/etc/hadoop$ mkdir -p /usr/local/hadoop_store/hdfs/namenode
mkdir: cannot create directory '/usr/local/hadoop_store': Permission denied
chhetank@chhetank-Lenovo-G560: /usr/local/hadoop/etc/hadoop$ chmod 644 /usr/local/hadoop_store/hdfs/namenode
chmod: cannot access '/usr/local/hadoop_store/hdfs/namenode': No such file or directory
chhetank@chhetank-Lenovo-G560: /usr/local/hadoop/etc/hadoop$ sudo mkdir -p /usr/local/hadoop_store/hdfs/namenode
chhetank@chhetank-Lenovo-G560: /usr/local/hadoop/etc/hadoop$ sudo mkdir -p /usr/local/hadoop_store/hdfs/datanode
chhetank@chhetank-Lenovo-G560: /usr/local/hadoop/etc/hadoop$
```

[Figure 14] Creation of namenode and datanode directories.

Now open /usr/local/hadoop/etc/hadoop/hdfs-site.xml and enter the following content in between the tag <configuration></configuration>

```
<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>
```



[Figure 15] Configuration of Namenode and Datanode

Now Make sure that dir has right owner and permission /usr/local/hadoop_store.

```

Sudo chown hduser:hadoop -R /usr/local/hadoop_store
sudo chmod 777 -R /usr/local/hadoop_store

```



[Figure 16] Checking permission and owner for datanode and namenode directory.

7. 7) Format Namenode

```

cd /usr/local/hadoop/
bin/hdfs namenode -format

```

[Figure 17] Formatting namenode method 1


```
chetank@chetank-Lenovo-G560: /usr/local/hadoop/bin
chetank@chetank-Lenovo-G560: /usr/local/hadoop/bin$ ./hdfs namenode -format
16/06/30 20:00:11 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = chetank-Lenovo-G560/127.0.1.1
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 2.7.0
STARTUP_MSG: classpath = /usr/local/hadoop/etc/hadoop:/usr/local/hadoop/share/hadoop/common/lib/curator-recipes-2.7.1.jar:/usr/local/hadoop/shar
e/hadoop/common/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/usr/local/hadoop/share/hadoop/common/lib/jettison-1.1.jar:/usr/local/hadoop/share/hadoo
p/common/lib/jersey-json-1.9.jar:/usr/local/hadoop/share/hadoop/common/lib/asm-3.2.jar:/usr/local/hadoop/share/hadoop/common/lib/jetty-6.1.26.jar:/
usr/local/hadoop/share/hadoop/common/lib/api-asn1-api-1.0.0-M20.jar:/usr/local/hadoop/share/hadoop/common/lib/hadoop-auth-2.7.0.jar:/usr/local/ha
doo/share/hadoop/common/lib/curator-framework-2.7.1.jar:/usr/local/hadoop/share/hadoop/common/lib/jackson-jaxrs-1.9.13.jar:/usr/local/hadoop/shar
e/hadoop/common/lib/jets3t-0.9.0.jar:/usr/local/hadoop/share/hadoop/common/lib/jackson-core-asl-1.9.13.jar:/usr/local/hadoop/share/hadoop/common/l
ib/zookeeper-3.4.6.jar:/usr/local/hadoop/share/hadoop/common/lib/servlet-api-2.5.jar:/usr/local/hadoop/share/hadoop/common/lib/mockito-all-1.8.5.j
ar:/usr/local/hadoop/share/hadoop/common/lib/commons-logging-1.1.3.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-beanutils-1.7.0.jar:/usr/
local/hadoop/share/hadoop/common/lib/curator-client-2.7.1.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-beanutils-core-1.8.0.jar:/u
sr/local/hadoop/share/hadoop/common/lib/commons-io-2.4.jar:/usr/local/hadoop/share/hadoop/common/lib/log4j-1.2.17.jar:/usr/local/hadoop/share/hadoop/common/lib/avro
-1.7.4.jar:/usr/local/hadoop/share/hadoop/common/lib/curator-0.52.jar:/usr/local/hadoop/share/hadoop/common/lib/jsr305-3.0.0.jar:/usr/local/hadoop/share/hadoop/co
mon/lib/netty-3.6.2.Final.jar:/usr/local/hadoop/share/hadoop/common/lib/jersey-server-1.9.jar:/usr/local/hadoop/share/hadoop/common/lib/jsp-api-2.1.jar:/usr/local/hadoop/share/ha
doo/common/lib/netty-3.6.2.Final.jar:/usr/local/hadoop/share/hadoop/common/lib/jsr305-3.0.0.jar:/usr/local/hadoop/share/hadoop/common/lib/commons
-digester-1.8.jar:/usr/local/hadoop/share/hadoop/common/lib/jersey-core-1.9.jar:/usr/local/hadoop/share/hadoop/common/lib/jackson-xc-1.9.13.jar:/u
sr/local/hadoop/share/hadoop/common/lib/commons-codec-1.4.jar:/usr/local/hadoop/share/hadoop/common/lib/guava-11.0.2.jar:/usr/local/hadoop/share/h
adoop/common/lib/jackson-mapper-asl-1.9.13.jar:/usr/local/hadoop/share/hadoop/common/lib/stax-api-1.0-2.jar:/usr/local/hadoop/share/hadoop/common/
lib/java-xmlbuilder-0.4.jar:/usr/local/hadoop/share/hadoop/common/lib/xz-1.0.jar:/usr/local/hadoop/share/hadoop/common/lib/jsch-0.1.42.jar:/usr/lo
cal/hadoop/share/hadoop/common/lib/commons-cli-1.2.jar:/usr/local/hadoop/share/hadoop/common/lib/httpclient-4.2.5.jar:/usr/local/hadoop/share/hado
o/common/lib/jetty-util-6.1.26.jar:/usr/local/hadoop/share/hadoop/common/lib/jaxb-api-2.2.2.jar:/usr/local/hadoop/share/hadoop/common/lib/gson-
2.4.jar:/usr/local/hadoop/share/hadoop/common/lib/hadoop-annotations-2.7.0.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-collections-3.2
.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-net-3.1.jar:/usr/local/hadoop/share/hadoop/common/lib/hamcrest-core-1.3.jar:/usr/local/ha
op/share/hadoop/common/lib/htrace-core-3.1.0-incubating.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-configuration-1.6.jar:/usr/local/ha
op/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar:/usr/local/hadoop/share/hadoop/common/lib/api-util-1.0.0-M20.jar:/usr/local/hadoop/share/hado
op/common/lib/commons-compress-1.4.1.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-math3-3.1.1.jar:/usr/local/hadoop/share/hadoop/common/l
ib/activation-1.1.jar:/usr/local/hadoop/share/hadoop/common/lib/snappy-java-1.0.4.1.jar:/usr/local/hadoop/share/hadoop/common/lib/slf4j-api-1.7.10
.jar:/usr/local/hadoop/share/hadoop/common/lib/junit-4.11.jar:/usr/local/hadoop/share/hadoop/hdfs:/usr/local/hadoop/share/hadoop/common/lib/protobuf
-java-2.5.0.jar:/usr/local/hadoop/share/hadoop/hdfs:/usr/local/hadoop/share/hadoop/common/lib/apacheds-118n-2.0.0-M15.jar:/usr/local/hadoop/share/hadoop/com
mon/lib/commons-httpclient-3.1.jar:/usr/local/hadoop/share/hadoop/common/lib/commons-lang-2.6.jar:/usr/local/hadoop/share/hadoop/common/lib/httpco
re-4.2.5.jar:/usr/local/hadoop/share/hadoop/common/hadoop-common-2.7.0.jar:/usr/local/hadoop/share/hadoop/common/hadoop-common-2.7.0-tests.jar:/u
sr/local/hadoop/share/hadoop/hdfs/lib/jetty-6.1.26.jar:/usr/local/hadoop/share/hadoop/hdfs/lib/jackson-core-asl-1.9.13.jar:/usr/local/hadoop/share
/hadoop/hdfs/lib/servlet-api-2.5.jar:/usr/local/hadoop/share/hadoop/hdfs/lib/commons-logging-1.1.3.jar:/usr/local/hadoop/share/hadoop/hdfs/lib/com
mons-io-2.4.jar:/usr/local/hadoop/share/hadoop/hdfs/lib/log4j-1.2.17.jar:/usr/local/hadoop/share/hadoop/hdfs/lib/xmlenc-0.52.jar:/usr/local/hadoop
/share/hadoop/hdfs/lib/jersey-server-1.9.jar:/usr/local/hadoop/share/hadoop/hdfs/lib/netty-3.6.2.Final.jar:/usr/local/hadoop/share/hadoop/hdfs/lib
```

```
hduser@chetank-Lenovo-G560: /usr/local/hadoop
16/06/30 20:27:07 INFO blockmanagement.BlockManager: maxNumBlocksToLog = 1000
16/06/30 20:27:07 INFO namenode.FSNamesystem: fsOwner = hduser (auth:SIMPLE)
16/06/30 20:27:07 INFO namenode.FSNamesystem: supergroup = supergroup
16/06/30 20:27:07 INFO namenode.FSNamesystem: isPermissionEnabled = true
16/06/30 20:27:07 INFO namenode.FSNamesystem: HA Enabled: false
16/06/30 20:27:07 INFO namenode.FSNamesystem: Append Enabled: true
16/06/30 20:27:08 INFO util.GSet: Computing capacity for map INodeMap
16/06/30 20:27:08 INFO util.GSet: VM type = 64-bit
16/06/30 20:27:08 INFO util.GSet: 1.0% max memory 889 MB = 8.9 MB
16/06/30 20:27:08 INFO util.GSet: capacity = 2^20 = 1048576 entries
16/06/30 20:27:08 INFO namenode.FSDirectory: ACLs enabled? false
16/06/30 20:27:08 INFO namenode.FSDirectory: XAttrs enabled? true
16/06/30 20:27:08 INFO namenode.FSDirectory: Maximum size of an attr: 16384
16/06/30 20:27:08 INFO namenode.NameNode: Caching file names occurring more than 10 times
16/06/30 20:27:08 INFO util.GSet: Computing capacity for map cachedBlocks
16/06/30 20:27:08 INFO util.GSet: VM type = 64-bit
16/06/30 20:27:08 INFO util.GSet: 0.25% max memory 889 MB = 2.2 MB
16/06/30 20:27:08 INFO util.GSet: capacity = 2^18 = 262144 entries
16/06/30 20:27:08 INFO namenode.FSNamesystem: dfs.namenode.safemode.threshold-pct = 0.9990000128746033
16/06/30 20:27:08 INFO namenode.FSNamesystem: dfs.namenode.safemode.min.datanodes = 0
16/06/30 20:27:08 INFO namenode.FSNamesystem: dfs.namenode.safemode.extension = 30000
16/06/30 20:27:08 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
16/06/30 20:27:08 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
16/06/30 20:27:08 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
16/06/30 20:27:08 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
16/06/30 20:27:08 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
16/06/30 20:27:08 INFO util.GSet: Computing capacity for map NameNodeRetryCache
16/06/30 20:27:08 INFO util.GSet: VM type = 64-bit
16/06/30 20:27:08 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
16/06/30 20:27:08 INFO util.GSet: capacity = 2^15 = 32768 entries
16/06/30 20:27:08 WARN conf.Configuration: bad conf file: element not <property>
16/06/30 20:27:08 WARN conf.Configuration: bad conf file: element not <property>
16/06/30 20:27:08 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1253563418-127.0.1.1-1467298628354
16/06/30 20:27:08 INFO common.Storage: Storage directory /usr/local/hadoop/store/hdfs/namenode has been successfully formatted.
16/06/30 20:27:08 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
16/06/30 20:27:08 INFO util.ExitUtil: Exiting with status 0
16/06/30 20:27:08 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at chetank-Lenovo-G560/127.0.1.1
*****/
hduser@chetank-Lenovo-G560: /usr/local/hadoop$ bin/hdfs namenode -format
```

[Figure 18] Formatting namenode method 2

7.8) Start all Hadoop daemons

```
cd /usr/local/hadoop/
sbin/start-dfs.sh
```

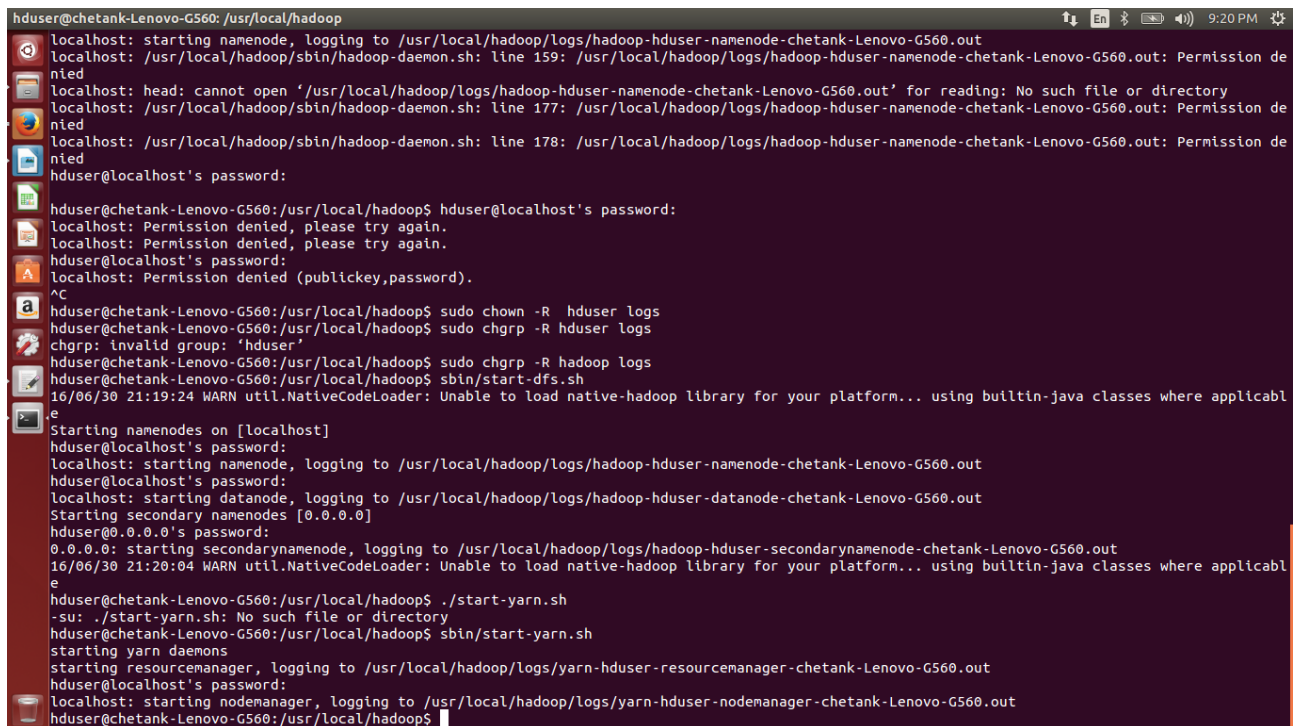


```
hduser@chetank-Lenovo-G560: /usr/local/hadoop
drwxr-xr-x 3 chetank chetank 4096 Apr 11 2015 etc
drwxr-xr-x 2 chetank chetank 4096 Apr 11 2015 bin
drwxr-xr-x 2 root root 4096 Jun 30 21:14 logs
hduser@chetank-Lenovo-G560:/usr/local/hadoop$ sudo chown -R logs
chown: missing operand after 'logs'
Try 'chown --help' for more information.
hduser@chetank-Lenovo-G560:/usr/local/hadoop$ sudo chown -R logs chetank
chown: invalid user: 'logs'
hduser@chetank-Lenovo-G560:/usr/local/hadoop$ sudo chown -R chetank logs
hduser@chetank-Lenovo-G560:/usr/local/hadoop$ ls -lrt
total 56
drwxr-xr-x 4 chetank chetank 4096 Apr 11 2015 share
drwxr-xr-x 2 chetank chetank 4096 Apr 11 2015 sbin
-rw-r--r-- 1 chetank chetank 1366 Apr 11 2015 README.txt
-rw-r--r-- 1 chetank chetank 101 Apr 11 2015 NOTICE.txt
-rw-r--r-- 1 chetank chetank 15429 Apr 11 2015 LICENSE.txt
drwxr-xr-x 2 chetank chetank 4096 Apr 11 2015 libexec
drwxr-xr-x 3 chetank chetank 4096 Apr 11 2015 lib
drwxr-xr-x 2 chetank chetank 4096 Apr 11 2015 include
drwxr-xr-x 3 chetank chetank 4096 Apr 11 2015 etc
drwxr-xr-x 2 chetank chetank 4096 Apr 11 2015 bin
drwxr-xr-x 2 chetank root 4096 Jun 30 21:14 logs
hduser@chetank-Lenovo-G560:/usr/local/hadoop$ sudo chgrp -R chetank logs
hduser@chetank-Lenovo-G560:/usr/local/hadoop$ ls -lrt
total 56
drwxr-xr-x 4 chetank chetank 4096 Apr 11 2015 share
drwxr-xr-x 2 chetank chetank 4096 Apr 11 2015 sbin
-rw-r--r-- 1 chetank chetank 1366 Apr 11 2015 README.txt
-rw-r--r-- 1 chetank chetank 101 Apr 11 2015 NOTICE.txt
-rw-r--r-- 1 chetank chetank 15429 Apr 11 2015 LICENSE.txt
drwxr-xr-x 2 chetank chetank 4096 Apr 11 2015 libexec
drwxr-xr-x 3 chetank chetank 4096 Apr 11 2015 lib
drwxr-xr-x 2 chetank chetank 4096 Apr 11 2015 include
drwxr-xr-x 3 chetank chetank 4096 Apr 11 2015 etc
drwxr-xr-x 2 chetank chetank 4096 Apr 11 2015 bin
drwxr-xr-x 2 chetank root 4096 Jun 30 21:14 logs
hduser@chetank-Lenovo-G560:/usr/local/hadoop$ sbin/start-dfs.sh
16/06/30 21:17:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicabl
e
Starting namenodes on [localhost]
hduser@localhost's password: 
```

[Figure 18] Starting hadoop daemon services - dfs

```
hduser@chetank-Lenovo-G560: /usr/local/hadoop
drwxr-xr-x 3 chetank chetank 4096 Apr 11 2015 etc
drwxr-xr-x 2 chetank chetank 4096 Apr 11 2015 bin
drwxr-xr-x 2 chetank chetank 4096 Jun 30 21:14 logs
hduser@chetank-Lenovo-G560:/usr/local/hadoop$ sbin/start-dfs.sh
16/06/30 21:17:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicabl
e
Starting namenodes on [localhost]
hduser@localhost's password:
localhost: chown: changing ownership of '/usr/local/hadoop/logs': Operation not permitted
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-namenode-chetank-Lenovo-G560.out
localhost: /usr/local/hadoop/sbin/hadoop-daemon.sh: line 159: /usr/local/hadoop/logs/hadoop-hduser-namenode-chetank-Lenovo-G560.out: Permission de
nied
localhost: head: cannot open '/usr/local/hadoop/logs/hadoop-hduser-namenode-chetank-Lenovo-G560.out' for reading: No such file or directory
localhost: /usr/local/hadoop/sbin/hadoop-daemon.sh: line 177: /usr/local/hadoop/logs/hadoop-hduser-namenode-chetank-Lenovo-G560.out: Permission de
nied
localhost: /usr/local/hadoop/sbin/hadoop-daemon.sh: line 178: /usr/local/hadoop/logs/hadoop-hduser-namenode-chetank-Lenovo-G560.out: Permission de
nied
hduser@localhost's password:
hduser@chetank-Lenovo-G560:/usr/local/hadoop$ hduser@localhost's password:
localhost: Permission denied, please try again.
localhost: Permission denied, please try again.
hduser@localhost's password:
localhost: Permission denied (publickey,password).
^C
hduser@chetank-Lenovo-G560:/usr/local/hadoop$ sudo chown -R hduser logs
hduser@chetank-Lenovo-G560:/usr/local/hadoop$ sudo chgrp -R hduser logs
chgrp: invalid group: 'hduser'
hduser@chetank-Lenovo-G560:/usr/local/hadoop$ sudo chgrp -R hadoop logs
hduser@chetank-Lenovo-G560:/usr/local/hadoop$ sbin/start-dfs.sh
16/06/30 21:19:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicabl
e
Starting namenodes on [localhost]
hduser@localhost's password:
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-namenode-chetank-Lenovo-G560.out
hduser@localhost's password:
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-datanode-chetank-Lenovo-G560.out
Starting secondary namenodes [0.0.0.0]
hduser@0.0.0.0's password:
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hduser-secondarynamenode-chetank-Lenovo-G560.out
```

7.9) sbin/start-yarn.sh



```
hduser@chetank-Lenovo-G560: /usr/local/hadoop
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-namenode-chetank-Lenovo-G560.out
localhost: /usr/local/hadoop/sbin/hadoop-daemon.sh: line 159: /usr/local/hadoop/logs/hadoop-hduser-namenode-chetank-Lenovo-G560.out: Permission de
nied
localhost: head: cannot open '/usr/local/hadoop/logs/hadoop-hduser-namenode-chetank-Lenovo-G560.out' for reading: No such file or directory
localhost: /usr/local/hadoop/sbin/hadoop-daemon.sh: line 177: /usr/local/hadoop/logs/hadoop-hduser-namenode-chetank-Lenovo-G560.out: Permission de
nied
localhost: /usr/local/hadoop/sbin/hadoop-daemon.sh: line 178: /usr/local/hadoop/logs/hadoop-hduser-namenode-chetank-Lenovo-G560.out: Permission de
nied
hduser@localhost's password:
hduser@chetank-Lenovo-G560: /usr/local/hadoop$ hduser@localhost's password:
localhost: Permission denied, please try again.
localhost: Permission denied, please try again.
hduser@localhost's password:
localhost: Permission denied (publickey,password).
^C
hduser@chetank-Lenovo-G560: /usr/local/hadoop$ sudo chown -R hduser logs
hduser@chetank-Lenovo-G560: /usr/local/hadoop$ sudo chgrp -R hduser logs
chgrp: invalid group: 'hduser'
hduser@chetank-Lenovo-G560: /usr/local/hadoop$ sudo chgrp -R hadoop logs
hduser@chetank-Lenovo-G560: /usr/local/hadoop$ sbin/start-dfs.sh
16/06/30 21:19:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicabl
e
Starting namenodes on [localhost]
hduser@localhost's password:
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-namenode-chetank-Lenovo-G560.out
hduser@localhost's password:
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-datanode-chetank-Lenovo-G560.out
Starting secondary namenodes [0.0.0.0]
hduser@0.0.0.0's password:
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hduser-secondarynamenode-chetank-Lenovo-G560.out
16/06/30 21:20:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicabl
e
hduser@chetank-Lenovo-G560: /usr/local/hadoop$ ./start-yarn.sh
-su: ./start-yarn.sh: No such file or directory
hduser@chetank-Lenovo-G560: /usr/local/hadoop$ sbin/start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resourcemanager-chetank-Lenovo-G560.out
hduser@localhost's password:
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduser-nodemanager-chetank-Lenovo-G560.out
hduser@chetank-Lenovo-G560: /usr/local/hadoop$
```

[Figure 19] Starting hadoop deamon services – yarn

Note: in single node cluster , there is no need to start yarn, but for learning purpose we can start yarn.

7.10) Check your hadoop Services

\$ jps

If everything goes well, you will see below matrix.

```
hduser@chetank-Lenovo-G560: /usr/local/hadoop
localhost: Permission denied, please try again.
localhost: Permission denied, please try again.
hduser@localhost's password:
localhost: Permission denied (publickey,password).
^C
hduser@chetank-Lenovo-G560: /usr/local/hadoop$ sudo chown -R hduser logs
hduser@chetank-Lenovo-G560: /usr/local/hadoop$ sudo chgrp -R hduser logs
chgrp: invalid group: 'hduser'
hduser@chetank-Lenovo-G560: /usr/local/hadoop$ sudo chgrp -R hadoop logs
hduser@chetank-Lenovo-G560: /usr/local/hadoop$ sbin/start-dfs.sh
16/06/30 21:19:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicabl
e
Starting namenodes on [localhost]
hduser@localhost's password:
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-namenode-chetank-Lenovo-G560.out
hduser@localhost's password:
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-datanode-chetank-Lenovo-G560.out
Starting secondary namenodes [0.0.0.0]
hduser@0.0.0.0's password:
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hduser-secondarynamenode-chetank-Lenovo-G560.out
16/06/30 21:20:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicabl
e
hduser@chetank-Lenovo-G560: /usr/local/hadoop$ ./start-yarn.sh
-su: ./start-yarn.sh: No such file or directory
hduser@chetank-Lenovo-G560: /usr/local/hadoop$ sbin/start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resourcemanager-chetank-Lenovo-G560.out
hduser@localhost's password:
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduser-nodemanager-chetank-Lenovo-G560.out
hduser@chetank-Lenovo-G560: /usr/local/hadoop$ hadoop fs -ls
16/06/30 21:20:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicabl
e
ls: '.': No such file or directory
hduser@chetank-Lenovo-G560: /usr/local/hadoop$ jps
11664 NameNode
12691 Jps
11819 DataNode
12508 NodeManager
12012 SecondaryNameNode
12191 ResourceManager
hduser@chetank-Lenovo-G560: /usr/local/hadoop$
```

[Figure 20] Apache Hadoop Services status

7.11) Hadoop Cluster Applications UI

For Hadoop Cluster Applications UI, to check running jobs.

Open: <http://localhost:8088>

it will redirect to <http://localhost:8088/cluster>

where you can see matrix of running jobs.

All Applications - Mozilla Firefox

All Applications

localhost:8088/cluster

hadoop

All Applications

Cluster

- About
- Nodes
- Node Labels
- Applications
- NEW
- NEW SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- FINISHED
- FAILED
- KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCoers Used	VCoers Total	VCoers Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes
0	0	0	0	0	0 B	8 GB	0 B	0	8	0	1	0	0

Scheduler Metrics

Scheduler Type		Scheduling Resource Type		Minimum Allocation		Maximum Allocation	
Capacity Scheduler		[MEMORY]		<memory:1024, vCores:1>		<memory:8192, vCores:1>	

Show 20 entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress
No data available in table									

Showing 0 to 0 of 0 entries

[Figure 21] Apache Hadoop Cluster Application UI

Errors might occurs:

1) while starting start-dfs.sh service, it might throws permission denied at namenode/datanode-something.log

Solution:

```
mkdir /usr/local/hadoop/logs  
sudo chown -R hduser /usr/local/hadoop/  
sudo chgrp -R hadoop /usr/local/hadoop/
```

Note: namenode and datanode directories should be created under **hduser** user and **hadoop** group.

Read more:

[1] HDFS Architecture

Ref. <https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html>

HDFS Access and API

[2] HDFS Commands Guide

Ref. <https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/HDFSCommands.html>

[3] JAVA API

Ref. <http://hadoop.apache.org/docs/current/api/org/apache/hadoop/fs/FileSystem.html>

[4] C/libhdfs

Ref. <https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/LibHdfs.html>

[5] WebHDFS API

Ref. <https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-hdfs/WebHDFS.html>