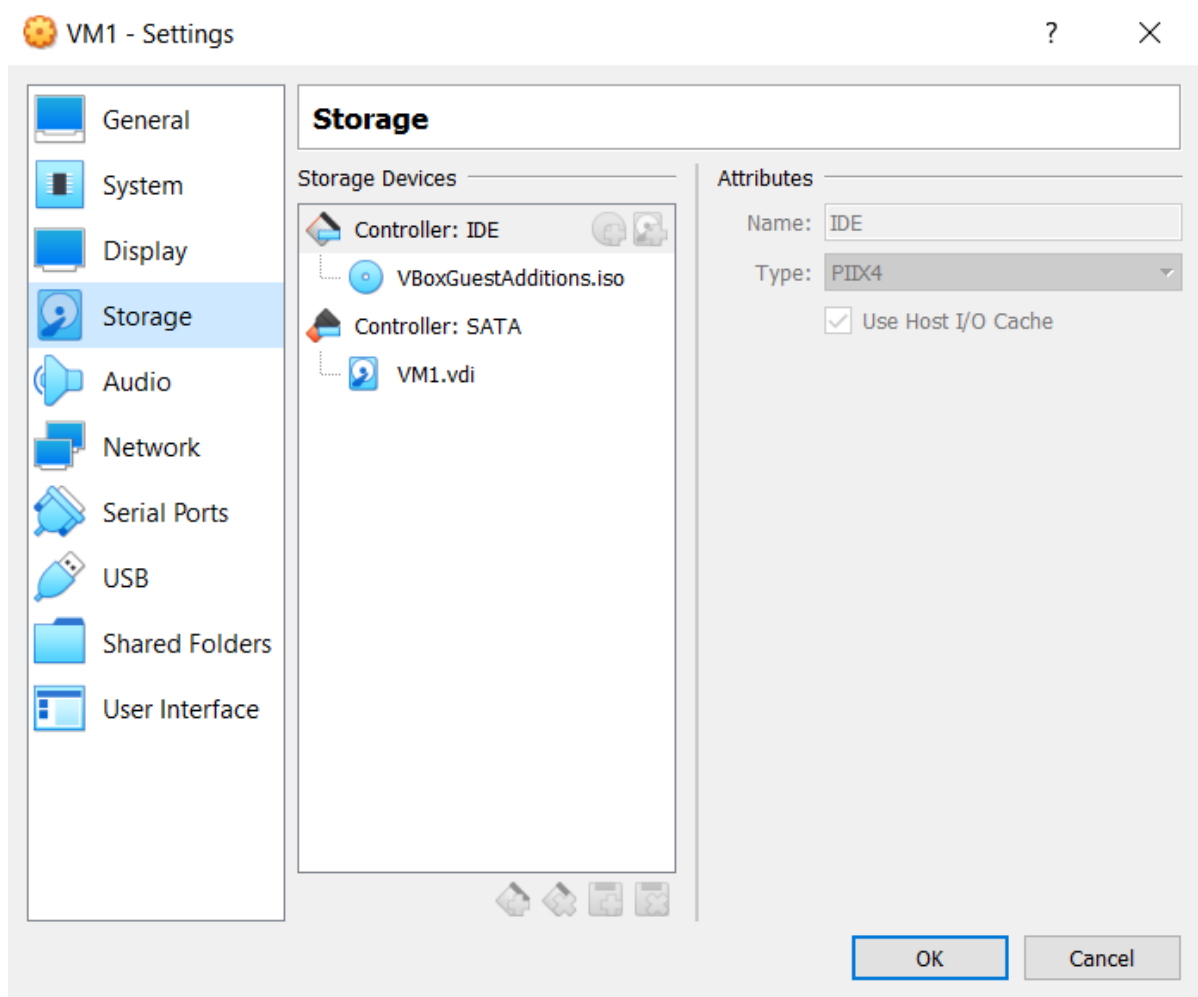
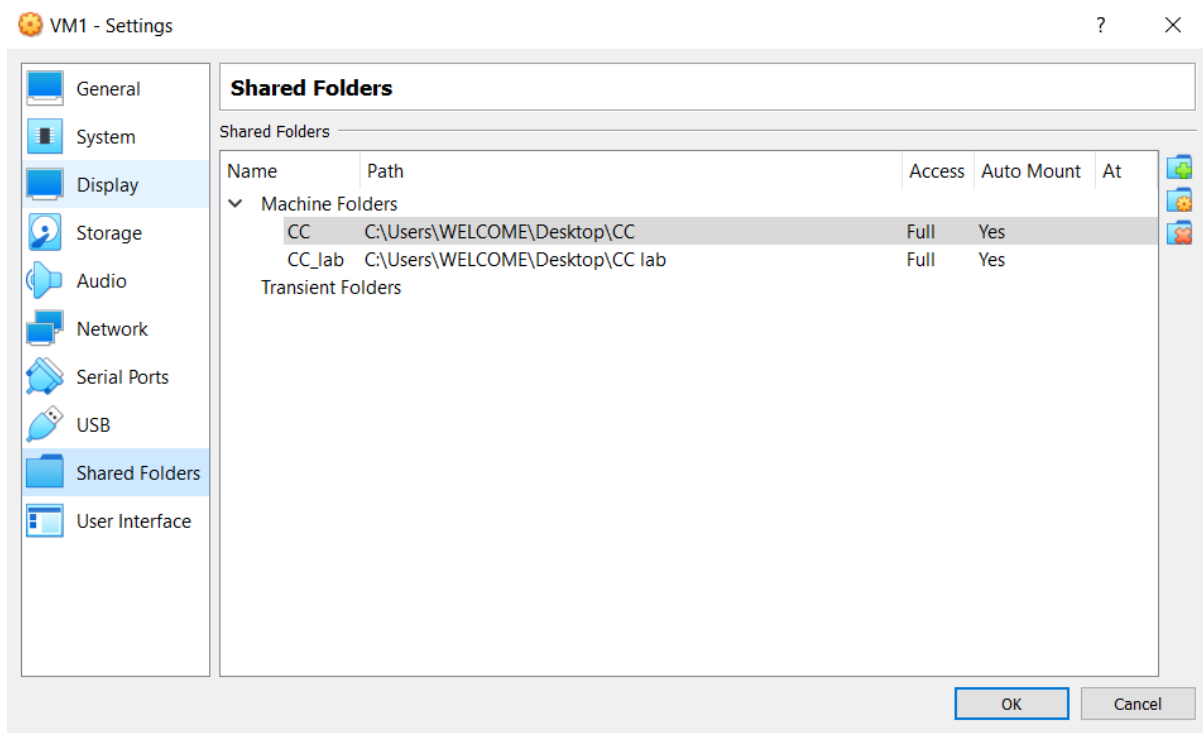
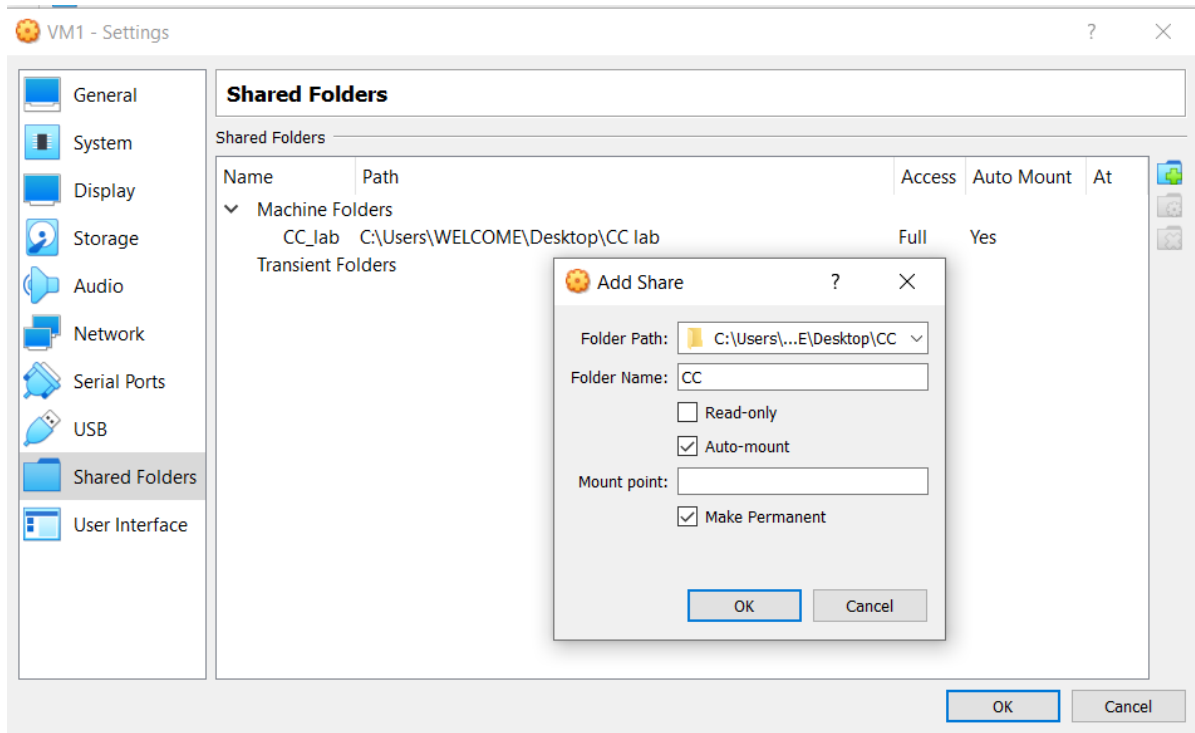


Ex.No.5 Installation of Single Node Hadoop and Executing Word Count Program

1) Adding VBox Guest Additions



```
hduser@vaishali-VirtualBox: ~  
hduser@vaishali-VirtualBox:~$ sudo adduser hduser vboxsf  
Adding user `hduser' to group `vboxsf' ...  
Adding user hduser to group vboxsf  
Done.  
hduser@vaishali-VirtualBox:~$
```



2) Update all the software in Ubuntu

```
vaishali@vaishali-VirtualBox:~$ sudo apt-get update
[sudo] password for vaishali:
Hit:1 http://in.archive.ubuntu.com/ubuntu xenial InRelease
Get:2 http://in.archive.ubuntu.com/ubuntu xenial-updates InRelease [109 kB]
Get:3 http://in.archive.ubuntu.com/ubuntu xenial-backports InRelease [107 kB]
Get:4 http://security.ubuntu.com/ubuntu xenial-security InRelease [109 kB]
Get:5 http://in.archive.ubuntu.com/ubuntu xenial-updates/main amd64 Packages [1,861 kB]
Get:6 http://in.archive.ubuntu.com/ubuntu xenial-updates/main i386 Packages [1,434 kB]
Get:7 http://security.ubuntu.com/ubuntu xenial-security/main amd64 Packages [1,456 kB]
Get:8 http://in.archive.ubuntu.com/ubuntu xenial-updates/main Translation-en [451 kB]
Get:9 http://in.archive.ubuntu.com/ubuntu xenial-updates/main amd64 DEP-11 Metadata [326 kB]
Get:10 http://in.archive.ubuntu.com/ubuntu xenial-updates/restricted amd64 Packages [10.2 kB]
Get:11 http://in.archive.ubuntu.com/ubuntu xenial-updates/restricted i386 Packages [10.2 kB]
Get:12 http://in.archive.ubuntu.com/ubuntu xenial-updates/universe amd64 Packages [1,178 kB]
Get:13 http://in.archive.ubuntu.com/ubuntu xenial-updates/universe i386 Packages [1,048 kB]
Get:14 http://in.archive.ubuntu.com/ubuntu xenial-updates/universe Translation-en [344 kB]
Get:15 http://in.archive.ubuntu.com/ubuntu xenial-updates/universe amd64 DEP-11 Metadata [279 kB]
Get:16 http://in.archive.ubuntu.com/ubuntu xenial-updates/universe DEP-11 64x64 Icons [419 kB]
Get:17 http://in.archive.ubuntu.com/ubuntu xenial-updates/multiverse amd64 Packages [23.0 kB]
Get:18 http://in.archive.ubuntu.com/ubuntu xenial-updates/multiverse i386 Packages [21.8 kB]
Get:19 http://in.archive.ubuntu.com/ubuntu xenial-updates/multiverse amd64 DEP-11 Metadata [5,960 B]
Get:20 http://in.archive.ubuntu.com/ubuntu xenial-backports/main amd64 Packages [9,812 B]
Get:21 http://in.archive.ubuntu.com/ubuntu xenial-backports/main i386 Packages [9,784 B]
Get:22 http://in.archive.ubuntu.com/ubuntu xenial-backports/main amd64 DEP-11 Metadata [3,332 B]
Get:23 http://in.archive.ubuntu.com/ubuntu xenial-backports/universe amd64 Packages [11.3 kB]
Get:24 http://in.archive.ubuntu.com/ubuntu xenial-backports/universe i386 Packages [10.9 kB]
Get:25 http://in.archive.ubuntu.com/ubuntu xenial-backports/universe amd64 DEP-11 Metadata [6,616 B]
Get:26 http://security.ubuntu.com/ubuntu xenial-security/main i386 Packages [1,064 kB]
Get:27 http://security.ubuntu.com/ubuntu xenial-security/main Translation-en [347 kB]
Get:28 http://security.ubuntu.com/ubuntu xenial-security/main amd64 DEP-11 Metadata [86.9 kB]
Get:29 http://security.ubuntu.com/ubuntu xenial-security/restricted amd64 Packages [9,824 B]
Get:30 http://security.ubuntu.com/ubuntu xenial-security/restricted i386 Packages [9,800 B]
Get:31 http://security.ubuntu.com/ubuntu xenial-security/universe amd64 Packages [754 kB]
Get:32 http://security.ubuntu.com/ubuntu xenial-security/universe i386 Packages [635 kB]
Get:33 http://security.ubuntu.com/ubuntu xenial-security/universe Translation-en [213 kB]
Get:34 http://security.ubuntu.com/ubuntu xenial-security/universe amd64 DEP-11 Metadata [128 kB]
Get:35 http://security.ubuntu.com/ubuntu xenial-security/universe DEP-11 64x64 Icons [193 kB]
Get:36 http://security.ubuntu.com/ubuntu xenial-security/multiverse amd64 Packages [8,228 B]
Get:37 http://security.ubuntu.com/ubuntu xenial-security/multiverse i386 Packages [8,480 B]
Get:38 http://security.ubuntu.com/ubuntu xenial-security/multiverse amd64 DEP-11 Metadata [2,464 B]
Fetched 12.7 MB in 14s (859 kB/s)
Reading package lists... Done
```

3) Create new user in system using the command **sudo adduser hduser**

```
vaishali@vaishali-VirtualBox:~$ sudo adduser hduser
Adding user `hduser' ...
Adding new group `hduser' (1001) ...
Adding new user `hduser' (1001) with group `hduser' ...
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
vaishali@vaishali-VirtualBox:~$ r
```

- 4) Add hduser to sudoers list using the command **sudo usermod -aG sudo hduser** and verify if the user belongs to Sudo Group using the command **groups hduser**.

```
vaishali@vaishali-VirtualBox:~$ sudo usermod -aG sudo hduser
vaishali@vaishali-VirtualBox:~$ groups hduser
hduser : hduser sudo
vaishali@vaishali-VirtualBox:~$
```

- 5) Logout and Login as hduser and install Java

```
hduser@vaishali-VirtualBox: ~
hduser@vaishali-VirtualBox:~$ java -version
java version "1.8.0_261"
Java(TM) SE Runtime Environment (build 1.8.0_261-b12)
Java HotSpot(TM) 64-Bit Server VM (build 25.261-b12, mixed mode)
```

- 6) Installing Secure Shell

a) sudo apt-get install ssh

```
hduser@vaishali-VirtualBox:~$ sudo apt-get install ssh
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  ncurses-term openssh-client openssh-server openssh-sftp-server ssh-import-id
Suggested packages:
  ssh-askpass libpam-ssh keychain monkeysphere rssh molly-guard
The following NEW packages will be installed:
  ncurses-term openssh-server openssh-sftp-server ssh ssh-import-id
The following packages will be upgraded:
  openssh-client
1 upgraded, 5 newly installed, 0 to remove and 611 not upgraded.
Need to get 640 kB/1,230 kB of archives.
After this operation, 5,248 kB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://in.archive.ubuntu.com/ubuntu xenial-updates/main amd64 openssh-sftp-server amd64 1:7.2p2-4ubuntu2.10 [38.8 kB]
Get:2 http://in.archive.ubuntu.com/ubuntu xenial-updates/main amd64 openssh-server amd64 1:7.2p2-4ubuntu2.10 [335 kB]
Get:3 http://in.archive.ubuntu.com/ubuntu xenial-updates/main amd64 ssh all 1:7.2p2-4ubuntu2.10 [7,072 B]
Get:4 http://in.archive.ubuntu.com/ubuntu xenial/main amd64 ncurses-term all 6.0+20160213-1ubuntu1 [249 kB]
Get:5 http://in.archive.ubuntu.com/ubuntu xenial/main amd64 ssh-import-id all 5.5-0ubuntu1 [10.2 kB]
Fetched 640 kB in 0s (1,213 kB/s)
Preconfiguring packages ...
(Reading database ... 176961 files and directories currently installed.)
Preparing to unpack .../openssh-client_1%3a7.2p2-4ubuntu2.10_amd64.deb ...
Unpacking openssh-client (1:7.2p2-4ubuntu2.10) over (1:7.2p2-4ubuntu2.2) ...
Selecting previously unselected package openssh-sftp-server.
Preparing to unpack .../openssh-sftp-server_1%3a7.2p2-4ubuntu2.10_amd64.deb ...
Unpacking openssh-sftp-server (1:7.2p2-4ubuntu2.10) ...
Selecting previously unselected package openssh-server.
Preparing to unpack .../openssh-server_1%3a7.2p2-4ubuntu2.10_amd64.deb ...
Unpacking openssh-server (1:7.2p2-4ubuntu2.10) ...
Selecting previously unselected package ssh.
Preparing to unpack .../ssh_1%3a7.2p2-4ubuntu2.10_all.deb ...
Unpacking ssh (1:7.2p2-4ubuntu2.10) ...
Selecting previously unselected package ncurses-term.
Preparing to unpack .../ncurses-term_6.0+20160213-1ubuntu1_all.deb ...
Unpacking ncurses-term (6.0+20160213-1ubuntu1) ...
Selecting previously unselected package ssh-import-id.
Preparing to unpack .../ssh-import-id_5.5-0ubuntu1_all.deb ...
Unpacking ssh-import-id (5.5-0ubuntu1) ...
Processing triggers for man-db (2.7.5-1) ...
```

```

Preconfiguring packages ...
(Reading database ... 176961 files and directories currently installed.)
Preparing to unpack .../openssh-client_1%3a7.2p2-4ubuntu2.10_amd64.deb ...
Unpacking openssh-client (1:7.2p2-4ubuntu2.10) over (1:7.2p2-4ubuntu2.2) ...
Selecting previously unselected package openssh-sftp-server.
Preparing to unpack .../openssh-sftp-server_1%3a7.2p2-4ubuntu2.10_amd64.deb ...
Unpacking openssh-sftp-server (1:7.2p2-4ubuntu2.10) ...
Selecting previously unselected package openssh-server.
Preparing to unpack .../openssh-server_1%3a7.2p2-4ubuntu2.10_amd64.deb ...
Unpacking openssh-server (1:7.2p2-4ubuntu2.10) ...
Selecting previously unselected package ssh.
Preparing to unpack .../ssh_1%3a7.2p2-4ubuntu2.10_all.deb ...
Unpacking ssh (1:7.2p2-4ubuntu2.10) ...
Selecting previously unselected package ncurses-term.
Preparing to unpack .../ncurses-term_6.0+20160213-1ubuntu1_all.deb ...
Unpacking ncurses-term (6.0+20160213-1ubuntu1) ...
Selecting previously unselected package ssh-import-id.
Preparing to unpack .../ssh-import-id_5.5-0ubuntu1_all.deb ...
Unpacking ssh-import-id (5.5-0ubuntu1) ...
Processing triggers for man-db (2.7.5-1) ...
Processing triggers for systemd (229-4ubuntu19) ...
Processing triggers for ureadahead (0.100.0-19) ...
Processing triggers for ufw (0.35-0ubuntu2) ...
Setting up openssh-client (1:7.2p2-4ubuntu2.10) ...
Setting up openssh-sftp-server (1:7.2p2-4ubuntu2.10) ...
Setting up openssh-server (1:7.2p2-4ubuntu2.10) ...
Creating SSH2 RSA key; this may take some time ...
2048 SHA256:We8t70ihVS2xiCqJ9EN0eilGW9zFVXjR6XvJXvN4ylU root@vaishali-VirtualBox (RSA)
Creating SSH2 DSA key; this may take some time ...
1024 SHA256:mWxJp0qB74ApV2nokUs0j0Ly104MYhmCTIkQnLIWzGM root@vaishali-VirtualBox (DSA)
Creating SSH2 ECDSA key; this may take some time ...
256 SHA256:SM23bIRCw3Jp43Kovq0IyFkXT+eqRkpedQf/X+xttSc root@vaishali-VirtualBox (ECDSA)
Creating SSH2 ED25519 key; this may take some time ...
256 SHA256:v4lPSMYW9vjpfy8qAl2X/3LffJCpbw9MAFUSRD/LYmk root@vaishali-VirtualBox (ED25519)
Setting up ssh (1:7.2p2-4ubuntu2.10) ...
Setting up ncurses-term (6.0+20160213-1ubuntu1) ...
Setting up ssh-import-id (5.5-0ubuntu1) ...
Processing triggers for systemd (229-4ubuntu19) ...
Processing triggers for ureadahead (0.100.0-19) ...
Processing triggers for ufw (0.35-0ubuntu2) ...
hduser@vaishali-VirtualBox:~$ █

```

b) sudo apt-get install rsync

```

hduser@vaishali-VirtualBox:~$ sudo apt-get install rsync
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages will be upgraded:
  rsync
1 upgraded, 0 newly installed, 0 to remove and 610 not upgraded.
Need to get 0 B/329 kB of archives.
After this operation, 12.3 kB of additional disk space will be used.
(Reading database ... 179724 files and directories currently installed.)
Preparing to unpack .../rsync_3.1.1-3ubuntu1.3_amd64.deb ...
Unpacking rsync (3.1.1-3ubuntu1.3) over (3.1.1-3ubuntu1) ...
Processing triggers for systemd (229-4ubuntu19) ...
Processing triggers for ureadahead (0.100.0-19) ...
Processing triggers for man-db (2.7.5-1) ...
Setting up rsync (3.1.1-3ubuntu1.3) ...
hduser@vaishali-VirtualBox:~$ █

```

c) sudo apt-get install openssh-server


```
hduser@vaishali-VirtualBox:~$ sudo apt-get install openssh-server
Reading package lists... Done
Building dependency tree
Reading state information... Done
openssh-server is already the newest version (1:7.2p2-4ubuntu2.10).
openssh-server set to manually installed.
0 upgraded, 0 newly installed, 0 to remove and 610 not upgraded.
hduser@vaishali-VirtualBox:~$ █
```

d) sudo apt-get install openssh-client

```
hduser@vaishali-VirtualBox:~$ sudo apt-get install openssh-client
Reading package lists... Done
Building dependency tree
Reading state information... Done
openssh-client is already the newest version (1:7.2p2-4ubuntu2.10).
0 upgraded, 0 newly installed, 0 to remove and 610 not upgraded.
hduser@vaishali-VirtualBox:~$
```

e) Generate key pair and add the public key to the authorized_keys.

ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa

```
hduser@vaishali-VirtualBox:~$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
Generating public/private rsa key pair.
Created directory '/home/hduser/.ssh'.
Your identification has been saved in /home/hduser/.ssh/id_rsa.
Your public key has been saved in /home/hduser/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:2CDwoGjZPT8rgdxuUwgCN3XMT/AT2MtKu+40ygEFlz8 hduser@vaishali-VirtualBox
The key's randomart image is:
+---[RSA 2048]---+
| . *.o+o+.      |
| oooBo.+..o.    |
| o+ o++..+o.    |
| . o.+..E++..   |
| .o =.BS        |
| .. = 0         |
| ..=00         |
| ..++..        |
| ooo           |
+---[SHA256]-----+
hduser@vaishali-VirtualBox:~$
```

f) **sudo cp ~/.ssh/id_rsa.pub ~/.ssh/authorized_keys** and verify SSH, using the command **ssh localhost**

```

hduser@vaishali-VirtualBox:~$ sudo cp ~/.ssh/id_rsa.pub ~/.ssh/authorized_keys
hduser@vaishali-VirtualBox:~$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:SM23bIRCw3Jp43Kovq0IyFkXT+eqRkpedQf/X+xttSc.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 16.04.3 LTS (GNU/Linux 4.10.0-28-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

623 packages can be updated.
467 updates are security updates.

New release '18.04.5 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

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.

hduser@vaishali-VirtualBox:~$ █

```

7) Installation of Hadoop 3.1.4

a) Download latest version of Apache Hadoop package

b) Extract and move the hadoop to an installation directory

```
]$ sudo tar xzf hadoop-3.1.4.tar.gz
```

```
]$ sudo mv hadoop-3.1.4 /usr/local/hadoop
```

c) Add the JAVA_HOME to hadoop-env.sh file

```
]$ sudo gedit /usr/local/hadoop/etc/hadoop/hadoop-env.sh
```

d) Locate java_home and set **JAVA_HOME=/usr/local/java**

e) Add the following lines to ~/.bashrc **]\$ sudo gedit ~/.bashrc**

```
#HADOOP VARIABLES START
```

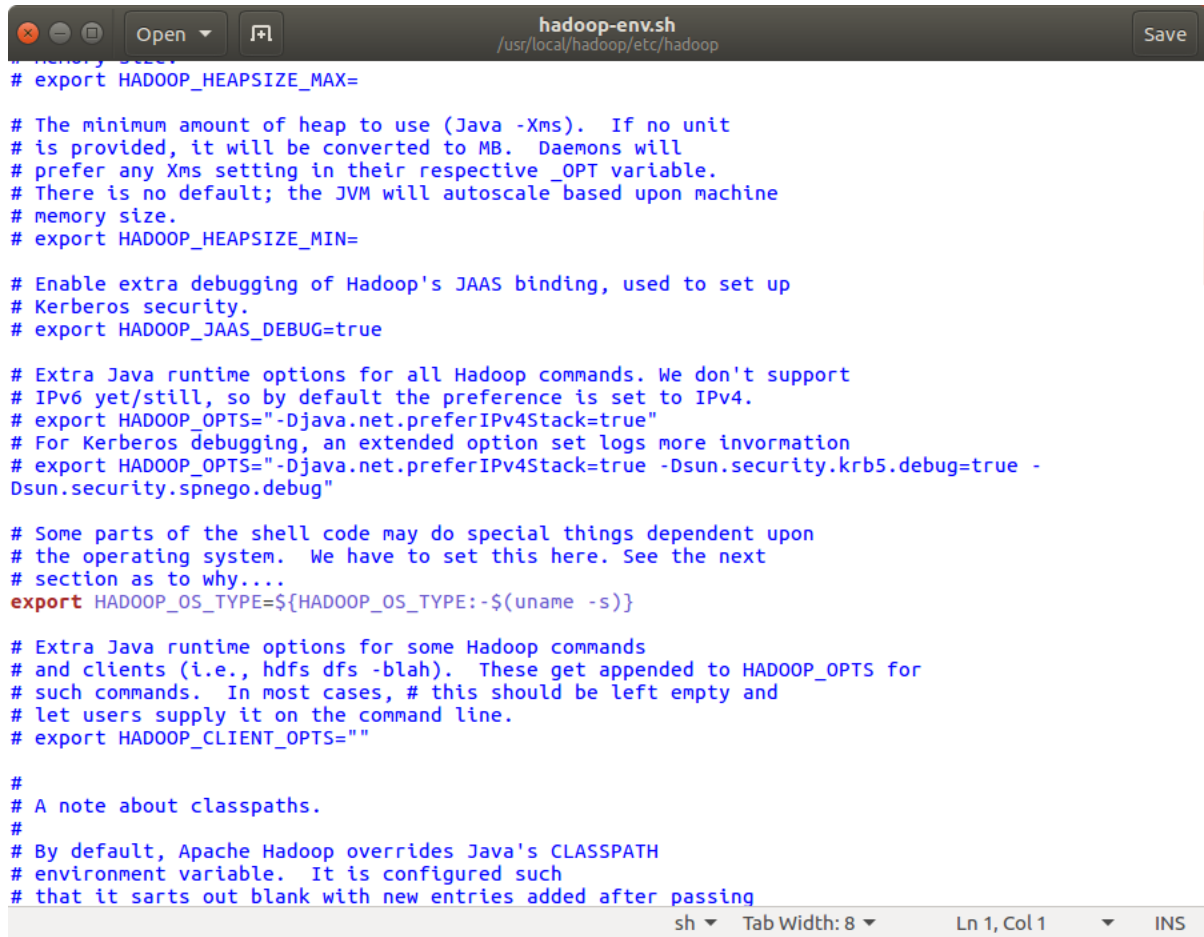
```

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"
export HADOOP_CLASSPATH=$JAVA_HOME/lib/tools.jar
```

```
#HADOOP VARIABLES END
```



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

# export HADOOP_HEAPSIZE_MAX=

# The minimum amount of heap to use (Java -Xms).  If no unit
# is provided, it will be converted to MB.  Daemons will
# prefer any Xms setting in their respective _OPT variable.
# There is no default; the JVM will autoscale based upon machine
# memory size.
# export HADOOP_HEAPSIZE_MIN=

# Enable extra debugging of Hadoop's JAAS binding, used to set up
# Kerberos security.
# export HADOOP_JAAS_DEBUG=true

# Extra Java runtime options for all Hadoop commands.  We don't support
# IPv6 yet/still, so by default the preference is set to IPv4.
# export HADOOP_OPTS="-Djava.net.preferIPv4Stack=true"
# For Kerberos debugging, an extended option set logs more information
# export HADOOP_OPTS="-Djava.net.preferIPv4Stack=true -Dsun.security.krb5.debug=true -
Dsun.security.spnego.debug"

# Some parts of the shell code may do special things dependent upon
# the operating system.  We have to set this here.  See the next
# section as to why....
export HADOOP_OS_TYPE=${HADOOP_OS_TYPE:-$(uname -s)}

# Extra Java runtime options for some Hadoop commands
# and clients (i.e., hdfs dfs -blah).  These get appended to HADOOP_OPTS for
# such commands.  In most cases, # this should be left empty and
# let users supply it on the command line.
# export HADOOP_CLIENT_OPTS=""

#
# A note about classpaths.
#
# By default, Apache Hadoop overrides Java's CLASSPATH
# environment variable.  It is configured such
# that it starts out blank with new entries added after passing
```

sh Tab Width: 8 Ln 1, Col 1 INS


```
.bashrc
# Add an "alert" alias for long running commands.  Use like so:
#   sleep 10; alert
alias alert='notify-send --urgency=low -i "${[ $? = 0 ]} && echo terminal || echo error"
"${history|tail -n1|sed -e '\''s/^s*[0-9]\+\s*//;s/[:&]\s*alert$/'\''}"

# 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

#HADOOP VARIABLES START
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"
export HADOOP_CLASSPATH=$JAVA_HOME/lib/tools.jar
#HADOOP VARIABLES END

sh Tab Width: 8 Ln 128, Col 61 INS
```

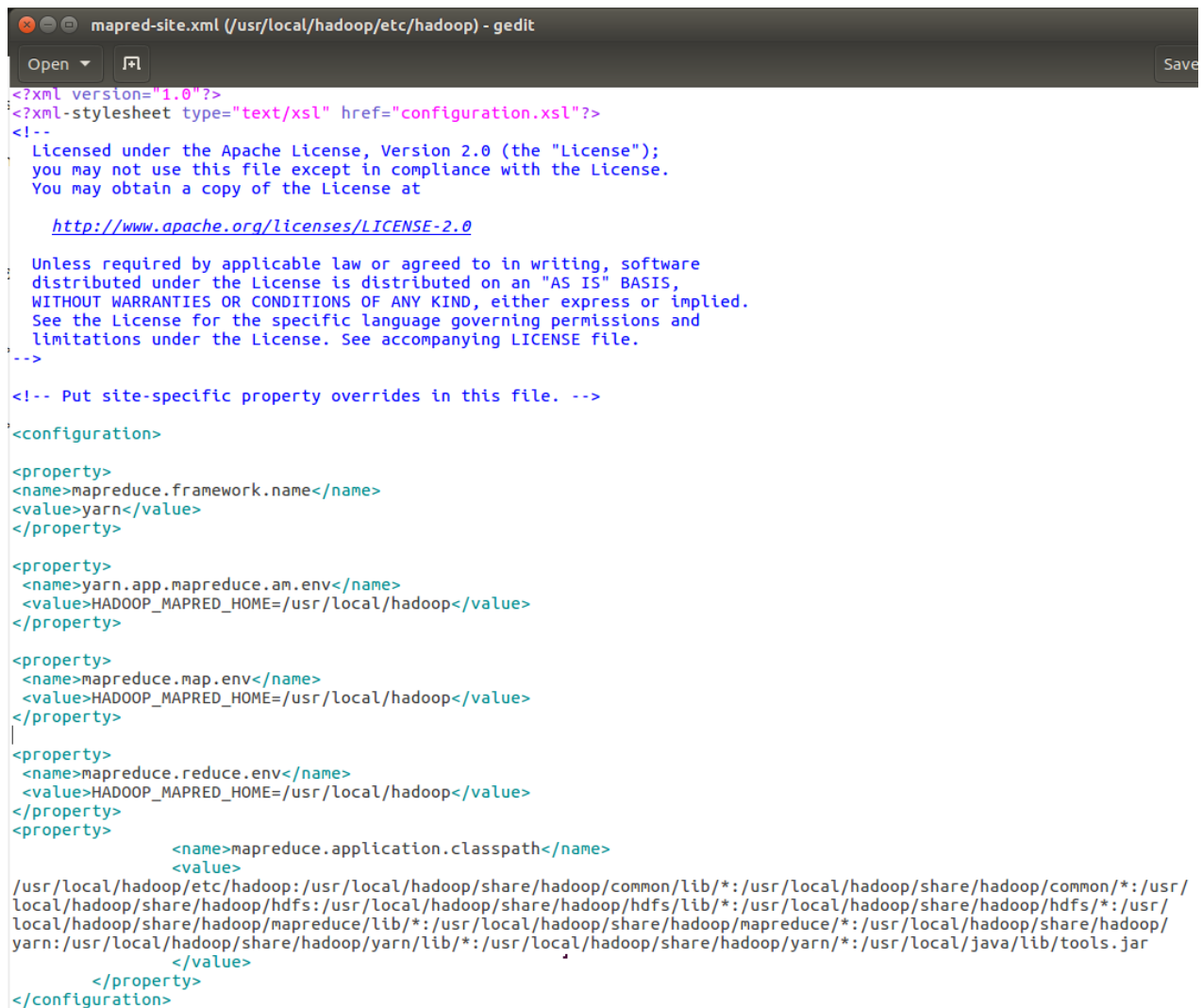
f) Refresh the bashrc file so that our environment variables can be accessed.

]\$ **source ~/.bashrc**

g) Check the Hadoop version

```
hduser@vaishali-VirtualBox:~$ hadoop version
Hadoop 3.1.4
Source code repository https://github.com/apache/hadoop.git -r 1e877761e8dadd71e
ffef30e592368f7fe66a61b
Compiled by gabota on 2020-07-21T08:05Z
Compiled with protoc 2.5.0
From source with checksum 38405c63945c88fdf7a6fe391494799b
This command was run using /usr/local/hadoop/share/hadoop/common/hadoop-common-3
.1.4.jar
hduser@vaishali-VirtualBox:~$ █
```

8) a) /usr/local/hadoop/etc/hadoop/mapred-site.xml



```
mapred-site.xml (/usr/local/hadoop/etc/hadoop) - gedit
Open [icon] Save

<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

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

  <property>
    <name>yarn.app.mapreduce.am.env</name>
    <value>HADOOP_MAPRED_HOME=/usr/local/hadoop</value>
  </property>

  <property>
    <name>mapreduce.map.env</name>
    <value>HADOOP_MAPRED_HOME=/usr/local/hadoop</value>
  </property>

  <property>
    <name>mapreduce.reduce.env</name>
    <value>HADOOP_MAPRED_HOME=/usr/local/hadoop</value>
  </property>

  <property>
    <name>mapreduce.application.classpath</name>
    <value>
/usr/local/hadoop/etc/hadoop:/usr/local/hadoop/share/hadoop/common/lib/*:/usr/local/hadoop/share/hadoop/common/*:/usr/
local/hadoop/share/hadoop/hdfs:/usr/local/hadoop/share/hadoop/hdfs/lib/*:/usr/local/hadoop/share/hadoop/hdfs/*:/usr/
local/hadoop/share/hadoop/mapreduce/lib/*:/usr/local/hadoop/share/hadoop/mapreduce/*:/usr/local/hadoop/share/hadoop/
yarn:/usr/local/hadoop/share/hadoop/yarn/lib/*:/usr/local/hadoop/share/hadoop/yarn/*:/usr/local/java/lib/tools.jar
    </value>
  </property>
</configuration>
```

b) Create Namenode and Datanode directories

```
j$ sudo mkdir -p /usr/local/hadoop_store/hdfs
```

```
hduser@vaishali-VirtualBox: ~$ sudo mkdir -p /usr/local/hadoop_store/hdfs
hduser@vaishali-VirtualBox: ~$ gedit /usr/local/hadoop/etc/hadoop/hdfs-site.xml

(gedit:3793): Gtk-WARNING **: Attempting to read the recently used resources file at '/home/hduser/.local/share/recently-used.xbel', but the parser failed: Failed to open file '/home/hduser/.local/share/recently-used.xbel': Permission denied.
hduser@vaishali-VirtualBox: ~$
```

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

```
hdfs-site.xml (/usr/local/hadoop/etc/hadoop) - gedit
Open [icon]
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

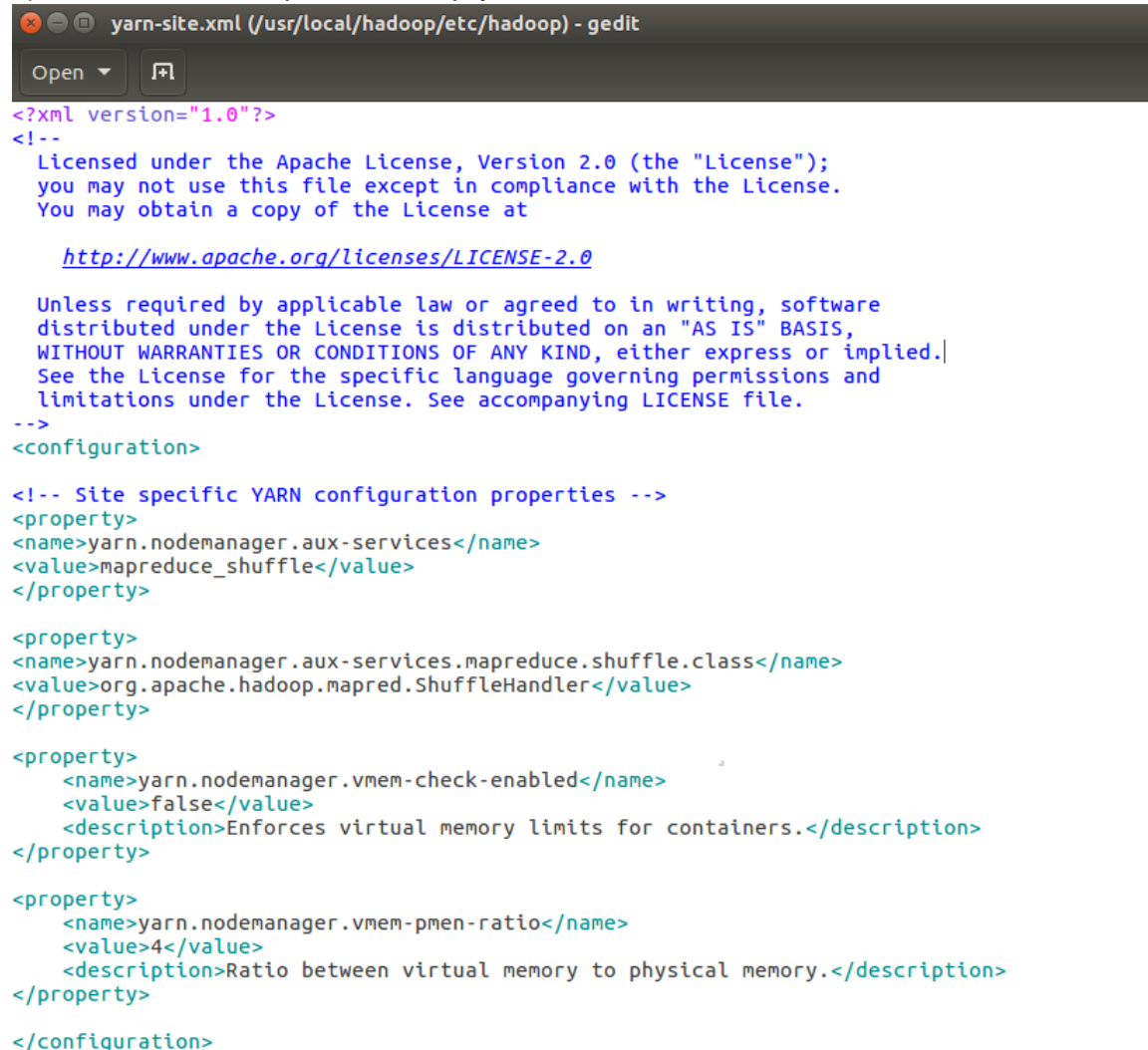
http://www.apache.org/licenses/LICENSE-2.0

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>dfs.replication</name>
<value>3</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>
</configuration>
```

c) /usr/local/hadoop/etc/hadoop/yarn-site.xml



```
<?xml version="1.0"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    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.
-->
<configuration>

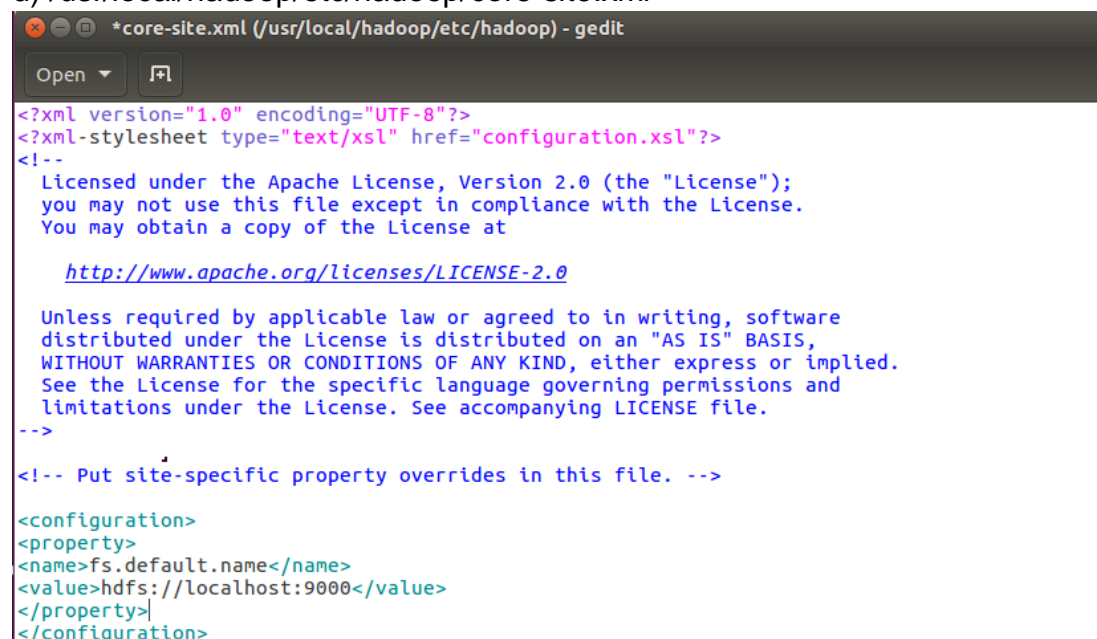
<!-- Site specific YARN configuration properties -->
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>

<property>
<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
<value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>

<property>
  <name>yarn.nodemanager.vmem-check-enabled</name>
  <value>>false</value>
  <description>Enforces virtual memory limits for containers.</description>
</property>

<property>
  <name>yarn.nodemanager.vmem-pmem-ratio</name>
  <value>4</value>
  <description>Ratio between virtual memory to physical memory.</description>
</property>
</configuration>
```

d) /usr/local/hadoop/etc/hadoop/core-site.xml



```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    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>
```

9) Change owner of `hadoop_store`

```
]$ sudo chown hduser:hduser -R /usr/local/Hadoop
```

```
]$ sudo chown hduser:hduser -R /usr/local/hadoop_store
```

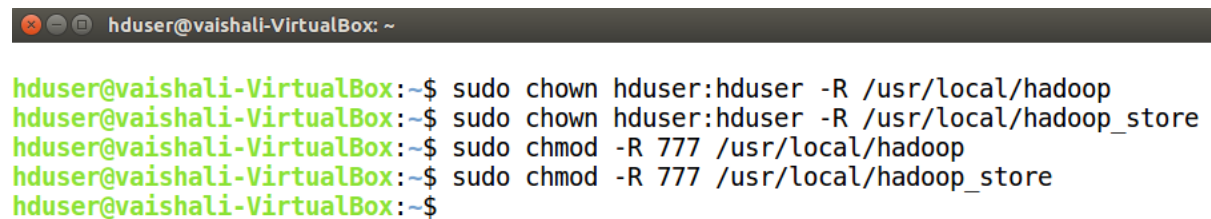
also give the folder the full permission

```
]$ sudo chmod -R 777 /usr/local/hadoop
```

```
]$ sudo chmod -R 777 /usr/local/hadoop_store
```

Format your HDFS, make sure you have logged in as `hadoop/hduser` user.

```
]$ hdfs namenode -format
```



```
hduser@vaishali-VirtualBox: ~  
hduser@vaishali-VirtualBox:~$ sudo chown hduser:hduser -R /usr/local/hadoop  
hduser@vaishali-VirtualBox:~$ sudo chown hduser:hduser -R /usr/local/hadoop_store  
hduser@vaishali-VirtualBox:~$ sudo chmod -R 777 /usr/local/hadoop  
hduser@vaishali-VirtualBox:~$ sudo chmod -R 777 /usr/local/hadoop_store  
hduser@vaishali-VirtualBox:~$
```

10) Start the hadoop dfs and yarn process

```
]$ start-all.sh
```

Check the hadoop deamons by running the following command

```
]$ jps
```



```
hduser@vaishali-VirtualBox: ~  
hduser@vaishali-VirtualBox:~$ start-all.sh  
WARNING: Attempting to start all Apache Hadoop daemons as hduser in 10 seconds.  
WARNING: This is not a recommended production deployment configuration.  
WARNING: Use CTRL-C to abort.  
Starting namenodes on [localhost]  
Starting datanodes  
Starting secondary namenodes [vaishali-VirtualBox]  
2020-10-07 10:23:59,033 WARN util.NativeCodeLoader: Unable to load native-hadoop  
library for your platform... using builtin-java classes where applicable  
Starting resourcemanager  
Starting nodemanagers  
hduser@vaishali-VirtualBox:~$ jps  
6311 ResourceManager  
6568 Jps  
5720 NameNode  
6010 SecondaryNameNode  
5836 DataNode  
6429 NodeManager  
hduser@vaishali-VirtualBox:~$
```

Resource Manager - localhost:8088

About

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

All Applications

localhost:8088/cluster

Search

Logged in as: drawho

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved
0	0	0	0	0	0 B	8 GB	0 B	0	8	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
1	0	0	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0

Showing 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCoers	Allocated Memory MB	Reserved CPU VCoers	Reserved Memory MB	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
No data available in table																				

Showing 0 to 0 of 0 entries

NameNode UI <http://localhost:9870/>

Namesnode information

All Applications

localhost:9870/dfshealth.html#tab-overview

Search

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

Overview 'localhost:9000' (active)

Started:	Tue Oct 06 23:54:15 +0530 2020
Version:	3.1.4, 1e877761e8dadd71effe30e592368f7fe66a61b
Compiled:	Tue Jul 21 13:35:00 +0530 2020 by gabota from branch-3.1.4
Cluster ID:	CID-fdc378a8-d436-44a4-9e2f-47f55061d209
Block Pool ID:	BP-1167002245-127.0.1.1-1602008320340

Summary

Security is off.
Safemode is off.
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
Heap Memory used 32.53 MB of 44.69 MB Heap Memory. Max Heap Memory is 483.38 MB.
Non Heap Memory used 49.16 MB of 50.52 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	8.73 GB
Configured Remote Capacity:	0 B
DFS Used:	28 KB (0%)
Non DFS Used:	7.06 GB
DFS Remaining:	1.21 GB (13.88%)
Block Pool Used:	28 KB (0%)

YARN UI <http://localhost:8042/>

ResourceManager

NodeManager

Node Information

List of Applications

List of Containers

Tools

http://localhost:8042/node

Search

ResourceManager information

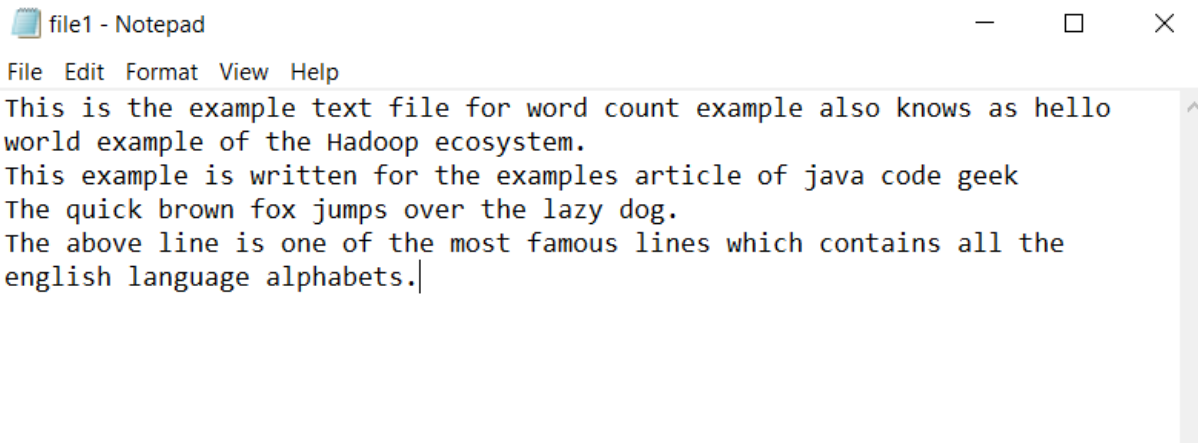
Total Vmem allocated for Containers	16.80 GB
Vmem enforcement enabled	true
Total Pmem allocated for Container	8 GB
Pmem enforcement enabled	true
Total VCoers allocated for Containers	8
Resource types	memory-mb (unit=Mi), vcores
NodeHealthyStatus	true
LastNodeHealthTime	Wed Oct 07 00:02:47 IST 2020
NodeHealthReport	
NodeManager started on	Tue Oct 06 23:54:41 IST 2020
NodeManager Version:	3.1.4 from 1e877761e8dadd71effe30e592368f7fe66a61b by gabota source checksum c3666d34f26916ba3cdf69aef06f5fbb on 2020-07-21T08:10Z
Hadoop Version:	3.1.4 from 1e877761e8dadd71effe30e592368f7fe66a61b by gabota source checksum 38405c63945c88df7a6fe391494799b on 2020-07-21T08:05Z

Execution of wordcount in Hadoop

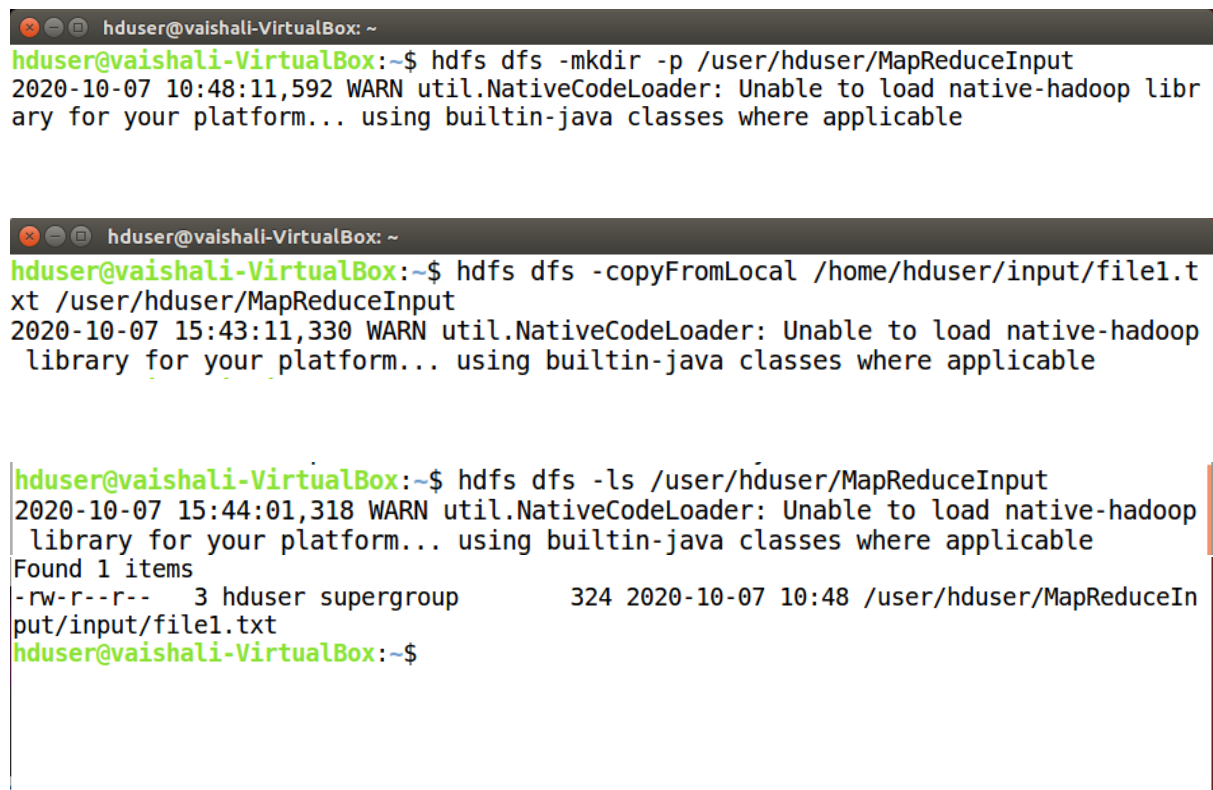
1. Managing files in HDFS.

Create an input file with few sentences in it.

File1.txt



Upload input file into HDFS using put command / copyFromLocal command.



```
hduser@vaishali-VirtualBox: ~  
hduser@vaishali-VirtualBox:~$ hdfs dfs -mkdir -p /user/hduser/MapReduceInput  
2020-10-07 10:48:11,592 WARN util.NativeCodeLoader: Unable to load native-hadoop lib  
ary for your platform... using builtin-java classes where applicable  
  
hduser@vaishali-VirtualBox:~$ hdfs dfs -copyFromLocal /home/hduser/input/file1.t  
xt /user/hduser/MapReduceInput  
2020-10-07 15:43:11,330 WARN util.NativeCodeLoader: Unable to load native-hadoop  
library for your platform... using builtin-java classes where applicable  
  
hduser@vaishali-VirtualBox:~$ hdfs dfs -ls /user/hduser/MapReduceInput  
2020-10-07 15:44:01,318 WARN util.NativeCodeLoader: Unable to load native-hadoop  
library for your platform... using builtin-java classes where applicable  
Found 1 items  
-rw-r--r--  3 hduser supergroup      324 2020-10-07 10:48 /user/hduser/MapReduceIn  
put/input/file1.txt  
hduser@vaishali-VirtualBox:~$
```

2. Write word count program in Java using Map and Reduce functions. Create word count program into a jar file using Eclipse. Execute the word count program's jar file using below command in hduser:

```
/usr/local/hadoop/bin/hadoop jar  
/usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-  
3.1.4.jar wordcount /user/hduser/MapReduceInput  
/user/hduser/MapReduceOutput
```

```
hduser@vaishali-VirtualBox:~$ /usr/local/hadoop/bin/hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.4.jar wordcount /user/hduser/MapReduceInput /user/hduser/MapReduceOutput
2020-10-07 23:40:41,545 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2020-10-07 23:40:42,958 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2020-10-07 23:40:44,520 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hduser/.staging/job_1602094229323_0001
2020-10-07 23:40:45,669 INFO input.FileInputFormat: Total input files to process : 1
2020-10-07 23:40:46,060 INFO mapreduce.JobSubmitter: number of splits:1
2020-10-07 23:40:46,416 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1602094229323_0001
2020-10-07 23:40:46,417 INFO mapreduce.JobSubmitter: Executing with tokens: []
2020-10-07 23:40:46,916 INFO conf.Configuration: resource-types.xml not found
2020-10-07 23:40:46,918 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2020-10-07 23:40:47,468 INFO impl.YarnClientImpl: Submitted application application_1602094229323_0001
2020-10-07 23:40:47,637 INFO mapreduce.Job: The url to track the job: http://vaishali-VirtualBox:8088/proxy/application_1602094229323_0001/
2020-10-07 23:40:47,638 INFO mapreduce.Job: Running job: job_1602094229323_0001
2020-10-07 23:41:03,472 INFO mapreduce.Job: Job job_1602094229323_0001 running in uber mode : false
2020-10-07 23:41:03,546 INFO mapreduce.Job: map 0% reduce 0%
2020-10-07 23:41:20,998 INFO mapreduce.Job: map 100% reduce 0%
2020-10-07 23:41:31,167 INFO mapreduce.Job: map 100% reduce 100%
2020-10-07 23:41:32,192 INFO mapreduce.Job: Job job_1602094229323_0001 completed successfully
2020-10-07 23:41:32,474 INFO mapreduce.Job: Counters: 53
```

File System Counters

```
FILE: Number of bytes read=517
FILE: Number of bytes written=446547
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=437
HDFS: Number of bytes written=339
HDFS: Number of read operations=8
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
```

Job Counters

```
Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=11073
Total time spent by all reduces in occupied slots (ms)=7583
Total time spent by all map tasks (ms)=11073
Total time spent by all reduce tasks (ms)=7583
Total vcore-milliseconds taken by all map tasks=11073
Total vcore-milliseconds taken by all reduce tasks=7583
Total megabyte-milliseconds taken by all map tasks=11338752
Total megabyte-milliseconds taken by all reduce tasks=7764992
```

```
total megabyte milliseconds taken by all reduce tasks=770
Map-Reduce Framework
  Map input records=4
  Map output records=58
  Map output bytes=554
  Map output materialized bytes=517
  Input split bytes=113
  Combine input records=58
  Combine output records=43
  Reduce input groups=43
  Reduce shuffle bytes=517
  Reduce input records=43
  Reduce output records=43
  Spilled Records=86
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=174
  CPU time spent (ms)=1490
  Physical memory (bytes) snapshot=329220096
  Virtual memory (bytes) snapshot=5201997824
  Total committed heap usage (bytes)=170004480
  Peak Map Physical memory (bytes)=214917120
  Peak Map Virtual memory (bytes)=2597552128
  Peak Reduce Physical memory (bytes)=114302976
  Peak Reduce Virtual memory (bytes)=2604445696
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=324
File Output Format Counters
  Bytes Written=339
```

```
hduser@vaishali-VirtualBox:~$ █
```

3) OUTPUT:

hdfs dfs -ls /user/hduser/MapReduceOutput

hdfs dfs -cat /user/hduser/MapReduceOutput/part-r-000000

```
hduser@vaishali-VirtualBox:~$ hdfs dfs -ls /user/hduser/MapReduceOutput
2020-10-07 23:49:05,061 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 hduser supergroup          0 2020-10-07 23:41 /user/hduser/MapReduceOutput/_SUCCESS
-rw-r--r-- 1 hduser supergroup    339 2020-10-07 23:41 /user/hduser/MapReduceOutput/part-r-000000
hduser@vaishali-VirtualBox:~$ hdfs dfs -cat /user/hduser/MapReduceOutput/part-r-000000
2020-10-07 23:50:03,707 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
using builtin-java classes where applicable
Hadoop 1
The 2
This 2
above 1
all 1
alphabets. 1
also 1
article 1
as 1
brown 1
code 1
contains 1
count 1
dog. 1
ecosystem. 1
english 1
example 4
examples 1
famous 1
file 1
for 2
fox 1
geek 1
hello 1
is 3
java 1
jumps 1
knows 1

knows 1
language 1
lazy 1
line 1
lines 1
most 1
of 3
one 1
over 1
quick 1
text 1
the 6
which 1
word 1
world 1
written 1
hduser@vaishali-VirtualBox:~$
```

4) Finally stop all Hadoop services.

```
hduser@vaishali-VirtualBox:~$ stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as hduser in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [vaishali-VirtualBox]
2020-10-08 00:06:53,383 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
using builtin-java classes where applicable
Stopping nodemanagers
localhost: WARNING: nodemanager did not stop gracefully after 5 seconds: Trying to kill with kill -9
Stopping resourcemanager
hduser@vaishali-VirtualBox:~$ jps
10964 Jps
hduser@vaishali-VirtualBox:~$ █
```