

1. Create instances on cloud:

(a) Create a lease.

The screenshot shows the 'Leases' creation form in the ChameleonCloud web interface. The 'Lease Name' field is empty. Below it, a note says: 'For leases shorter than 24 hours, use a lease length of zero days.' The 'Start Date' is set to 'Today', 'Lease Length (days)' is '1', and 'Ends' is 'Tomorrow'. The 'Start Time' is 'Now' and 'End Time' is 'Same time as now'. Under 'Physical Hosts', the 'Reserve Physical Host' checkbox is checked. In the 'Network' section, the 'Reserve Network' checkbox is unchecked. The 'Network Name' field is empty, and the 'Network Description' field is also empty. On the right, a table lists various nodes with their status (e.g., 'Normal', 'Degraded') and actions like 'Delete Lease'. At the bottom right of the form are 'Cancel' and 'Create' buttons.

(b) Set up instance name, select your lease, and choose instance number.

The screenshot shows the 'Launch Instance' form in the ChameleonCloud web interface. The 'Details' tab is selected. It asks for an 'Instance Name' (which is empty), a 'Description' (empty), and a 'Reservation' (empty). The 'Count' is set to '1'. To the right, usage statistics are shown: 'Total Instances (No Limit)', '15 Current Usage', and '1 Added'. On the left, a sidebar lists other configuration options: Source, Flavor, Networks, Network Ports, Security Groups, Key Pair, Configuration, Server Groups, Scheduler Hints, and Metadata. At the bottom are 'Cancel', 'Back', 'Next', and 'Launch Instance' buttons.

(c) Choose an image. It can be a new operating system, or a snapshot.

The screenshot shows the ChameleonCloud Instances management interface. On the left, a sidebar lists categories: Server, Network, Orchestration, Object Store, Reservations, and Identity. The main area has a header 'Details' and a sub-header 'Source *'. A dropdown menu 'Select Boot Source' is open, showing 'Flavor *' set to 'Image'. Below this, sections for 'Allocated' and 'Available' resources are shown. The 'Allocated' section lists 'Network Ports', 'Security Groups', and 'Key Pair'. The 'Available' section lists 'Configuration', 'Server Groups', 'Scheduler Hints', and 'Metadata'. Under 'Metadata', there are four entries: 'ARM64-CC-Ubuntu16.04' (updated 2/8/18, 633.19 MB, qcow2, Public), 'bigstack-cygnus-compute1' (updated 3/13/18, 1.07 GB, qcow2, Private), 'bigstack-cygnus-compute2' (updated 3/13/18, 1.07 GB, qcow2, Private), and 'bigstack-cygnus-head' (updated 3/13/18, 1.28 GB, qcow2, Private). On the right, there's a vertical 'Actions' column with buttons for 'Attach Interface' (repeated four times) and 'Edit Instance'. The bottom status bar shows the date and time: 8:06 PM, 5/1/2019.

(d) Server type, there is one choice.

The screenshot shows the 'Launch Instance' dialog box from ChameleonCloud. The sidebar on the left is identical to the main Instances page. The dialog has a header 'Launch Instance' and a sub-header 'Flavor'. A descriptive text block says: 'Flavors manage the sizing for the compute, memory and storage capacity of the instance.' Below this is a table for 'Allocated' resources, showing 'baremetal' with details: 'Hardware details for your instance can be found in our [resource discovery interface](#)'. A 'Public' checkbox is checked. The 'Available' section is currently empty. The right side of the dialog contains a vertical 'Actions' column with 'Attach Interface' and 'Edit Instance' buttons. At the bottom, there are 'Cancel', 'Back', 'Next >', and a large blue 'Launch Instance' button. The bottom status bar shows the date and time: 8:07 PM, 5/1/2019.

(e) Network. You can create your own network and use it, or use shared network (already created by other people).

The screenshot shows the ChameleonCloud Instances interface. On the left, a sidebar lists navigation options: Server, Network, Orchestration, Object Store, Reservations, and Identity. The main content area is titled 'Instances - ChameleonCloud' and shows the 'Instances' tab selected. Under the 'Network' section, the 'Networks' tab is active. It displays two tables: 'Allocated' and 'Available'. The 'Allocated' table has one item: 'sharednet1' associated with 'sharednet1-subnet', marked as 'Yes' for Shared, 'Up' for Admin State, and 'Active' for Status. The 'Available' table lists several networks: 'bn_exp_net' (associated with 'bn_exp_subnet', 'No' Shared, 'Up' Admin State, 'Active' Status), 'apxs' (associated with 'apsub', 'No' Shared, 'Up' Admin State, 'Active' Status), 'meta-openstack' (associated with 'meta-openstack-subnet1', 'No' Shared, 'Up' Admin State, 'Active' Status), 'xsearch.net' (associated with 'xsearch.subnet', 'No' Shared, 'Up' Admin State, 'Active' Status), 'DIC-Tests' (associated with 'DIC-Tests', 'No' Shared, 'Up' Admin State, 'Active' Status), 'meta-openstack-dhcp' (associated with 'meta-openstack-subnet2-dhcp', 'No' Shared, 'Up' Admin State, 'Active' Status), and 'aballmer-hw8-iscsi-net' (associated with 'isci-subnet', 'No' Shared, 'Up' Admin State, 'Active' Status). A search bar and filter button are present above the 'Available' table. On the right, there is an 'Actions' panel with multiple 'Attach Interface' buttons and an 'Edit Instance' button.

(f) Upload or paste your public keys to remote server. Each machine has different public keys. Server won't allow connection (for example, ssh) if it doesn't know your public key.

The screenshot shows the 'Launch Instance' wizard step. The sidebar on the left includes 'Server', 'Network', 'Orchestration', 'Object Store', 'Reservations', and 'Identity' sections. The 'Key Pair' tab is currently selected. The main area contains a 'Details' section with a note about key pairs for SSH access, a 'Source' section with '+ Create Key Pair' and 'Import Key Pair' buttons, and an 'Allocated' section showing one key pair named 'llu25_keys' with fingerprint 'f3:81:0e:9a:cf:0e:f8:e8:d2:6e:2d:d4:5e:d1:1a:4f'. Below this is an 'Available' section with a search bar and filter button, showing no items. At the bottom, there are 'Cancel', 'Back', 'Next >', and a large blue 'Launch Instance' button. The status bar at the bottom right shows the time as 8:11 PM and the date as 5/1/2019.

(g) Launch the instance. It'll take about half an hour.

(h) Generate float IP and associate it to your instance.

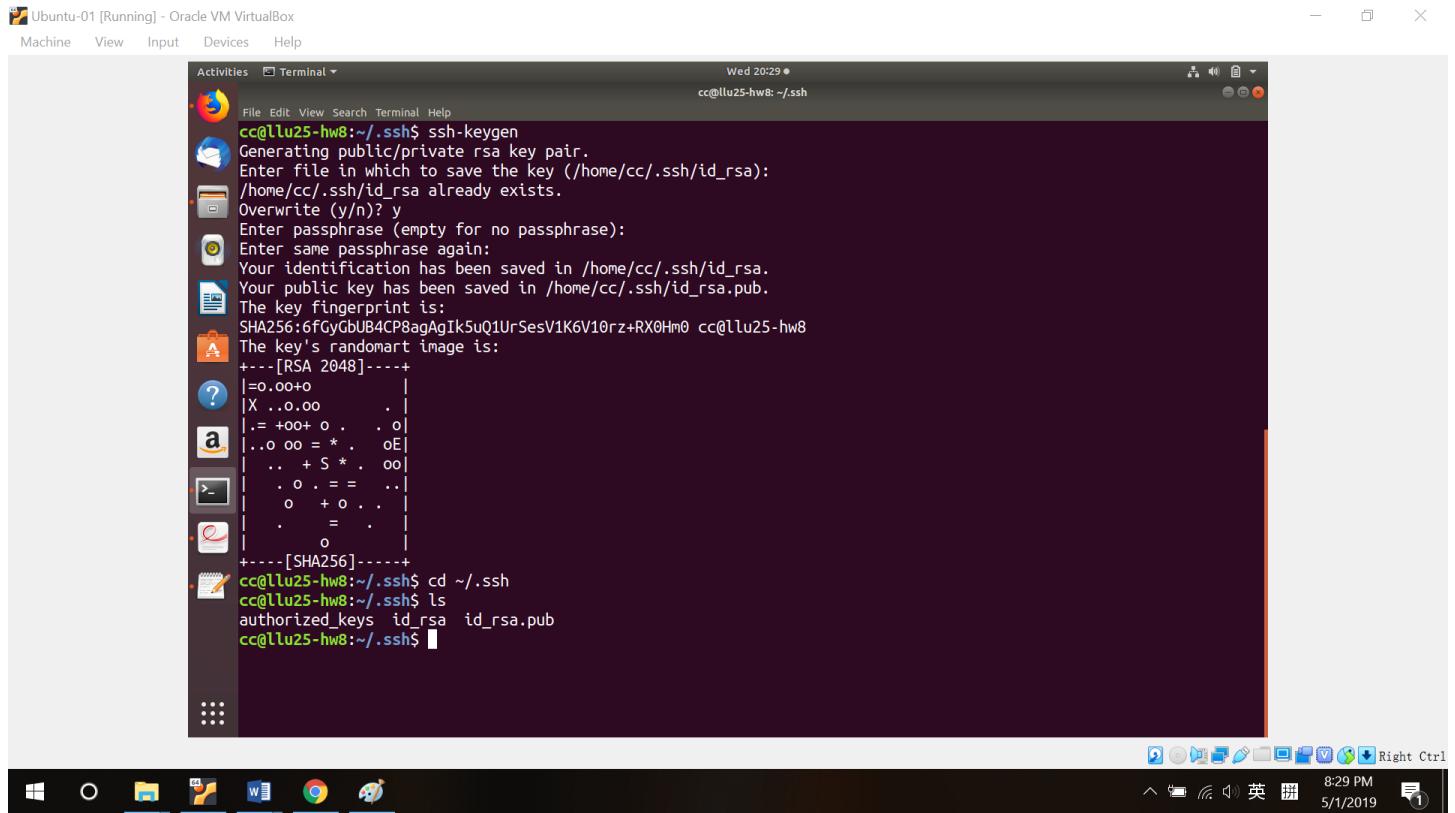
The screenshot shows the ChameleonCloud interface for managing floating IPs. The top navigation bar includes tabs for 'Floating IPs - ChameleonCloud' and a '+' button. The URL is https://chi.tacc.chameleoncloud.org/project/floating_ips/. The header features the Chameleon logo and the project ID CH-819437 • CHI@TACC. A user profile icon for 'liu25' is in the top right. Below the header, there's a search bar with 'Floating IP Address = ▾' and a dropdown, followed by 'Filter', 'Allocate IP To Project', and 'Release Floating IPs' buttons. The main content area has a 'Network' dropdown menu open, showing 'Network Topology' and 'Displaying 12 items'. Under 'Network Topology', there are sections for 'Networks', 'Routers', 'Security Groups', and 'Floating IPs'. The 'Floating IPs' section is highlighted with a blue background and contains a table with columns: IP Address, Description, Mapped Fixed IP Address, Pool, Status, and Actions. The table lists several entries, each with a checkbox next to the IP address and a 'Disassociate' button. The bottom of the screen shows a Windows taskbar with icons for File Explorer, Task View, Start, Task Manager, and a search bar. The system tray shows the date (5/1/2019), time (8:16 PM), battery level, signal strength, and network status.

(i) After above configuration, you will successfully create an instance with an IP address.

The screenshot shows the ChameleonCloud interface for managing instances. The top navigation bar includes tabs for 'Instances - ChameleonCloud' and a '+' button. The URL is https://chi.tacc.chameleoncloud.org/project/instances/. The header features the Chameleon logo and the project ID CH-819437 • CHI@TACC. A user profile icon for 'liu25' is in the top right. Below the header, there's a search bar with 'CH-819437 • CHI@TACC' and a dropdown, followed by 'Filter', 'Allocate IP To Project', and 'Release Floating IPs' buttons. The main content area has a 'Images' dropdown menu open, showing 'Displaying 15 items'. Under 'Images', there are sections for 'Key Pairs', 'Server Groups', and 'Instances'. The 'Instances' section is highlighted with a blue background and contains a table with columns: Instance Name, Image Name, IP Address, Flavor, Key Pair, Status, Availability Zone, Task, Power State, Time since created, and Actions. The table lists several instances, each with a checkbox next to the instance name and an 'Actions' button. The bottom of the screen shows a Windows taskbar with icons for File Explorer, Task View, Start, Task Manager, and a search bar. The system tray shows the date (5/1/2019), time (8:21 PM), battery level, signal strength, and network status.

2: SSH to your server

(a) Before copying your public key to server, you must generate it. Use “ssh-keygen” command to generate a key pair. If you didn’t change default path, the key pair should in `~/.ssh` directory.



The screenshot shows a terminal window in an Oracle VM VirtualBox environment. The window title is "Ubuntu-01 [Running] - Oracle VM VirtualBox". The terminal command history is as follows:

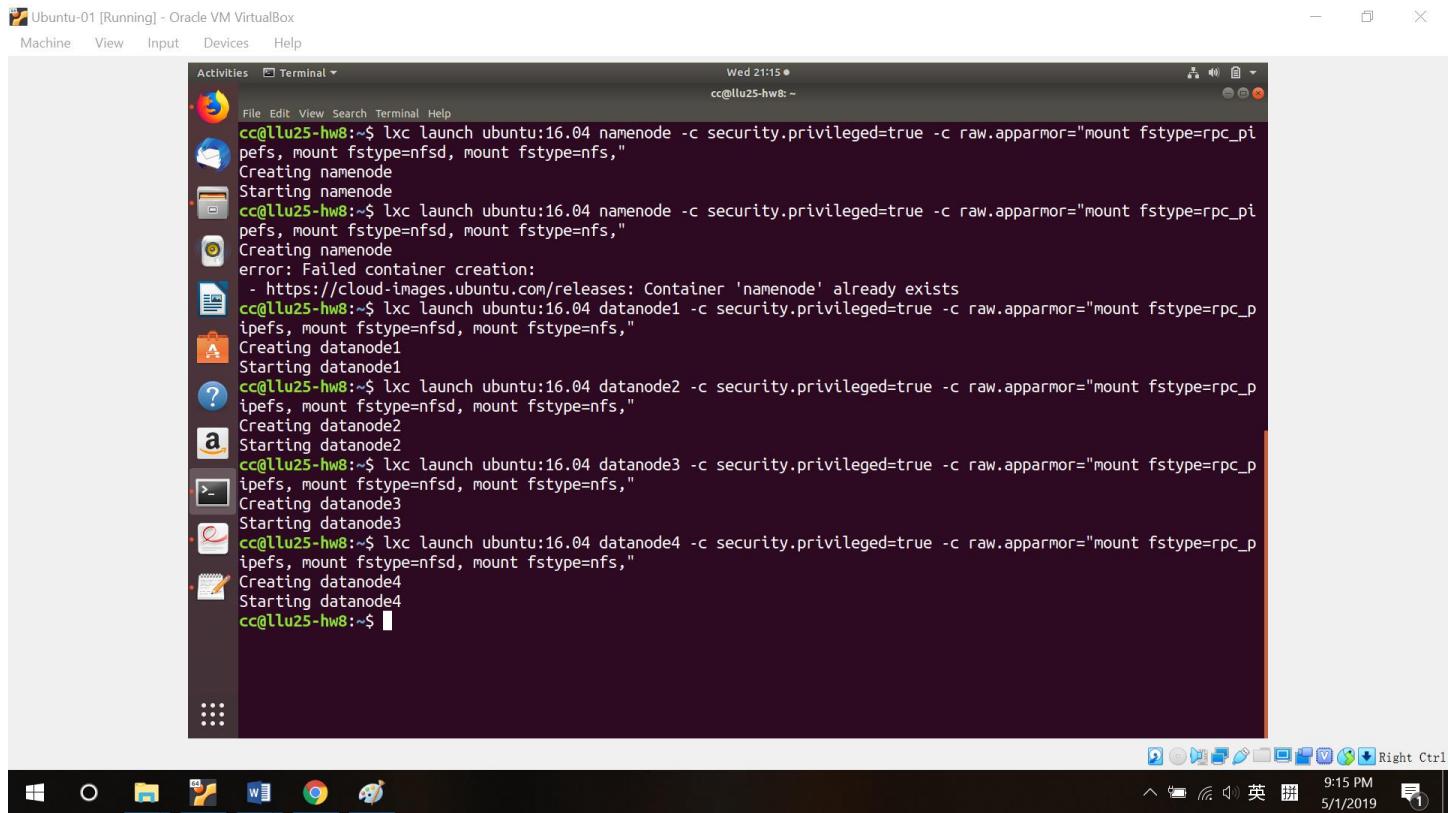
```
cc@llu25-hw8:~/ssh$ ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/home/cc/.ssh/id_rsa):
/home/cc/.ssh/id_rsa already exists.
Overwrite (y/n)? y
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/cc/.ssh/id_rsa.
Your public key has been saved in /home/cc/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:6fGyGbUB4CP8agAgIk5uQ1UrSesV1K6V10rz+RX0Hm0 cc@llu25-hw8
The key's randomart image is:
+---[RSA 2048]---+
|o..oo+o |
|X ..oo . |
|.= +oo+ o . oE|
|..o oo = * . oo|
|... + S * . oo|
| . o . = = ..|
| . o + o . |
| . = . |
| . o |
+---[SHA256]---+
cc@llu25-hw8:~/ssh$ cd ~./ssh
cc@llu25-hw8:~/ssh$ ls
authorized_keys id_rsa id_rsa.pub
cc@llu25-hw8:~/ssh$
```

The desktop taskbar at the bottom shows various application icons, and the system tray indicates the date and time as 5/1/2019 8:29 PM.

(b) You can copy your public key to a remote server by using “`ssh-copy-id user_name@server_ip`”.

3: Create and instance via lxc

(a) First install lxc if your system doesn't have. Then do some basic configurations like install a software to manage your file, set up disk capacity, set up bridged network, user permission and so on. All of them can be found on this website: <https://tutorials.ubuntu.com/tutorial/tutorial-setting-up-lxd-1604#0>. Then create your name node and server data nodes. In order to allow nfs services, you must tell your apparmor that such services is legal. Without such setting, you cannot use nfs. This is the most critical part!! We spend several days to figure it out. The solution sounds simple, just add 'fstype=nfs', but it's very very very critical!!!!

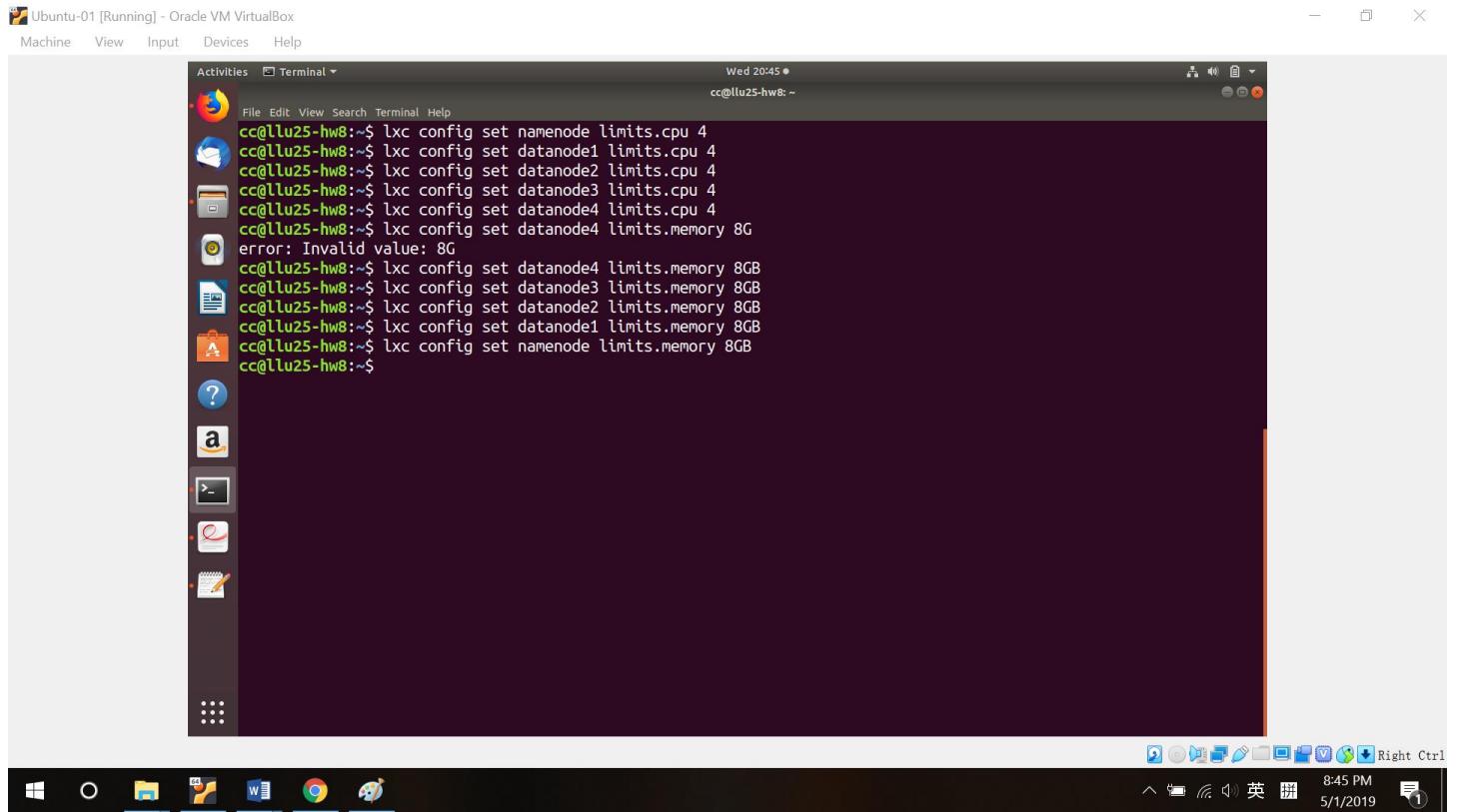


The screenshot shows a Linux desktop environment with a terminal window open. The terminal window title is "Ubuntu-01 [Running] - Oracle VM VirtualBox". The terminal content shows the following command execution:

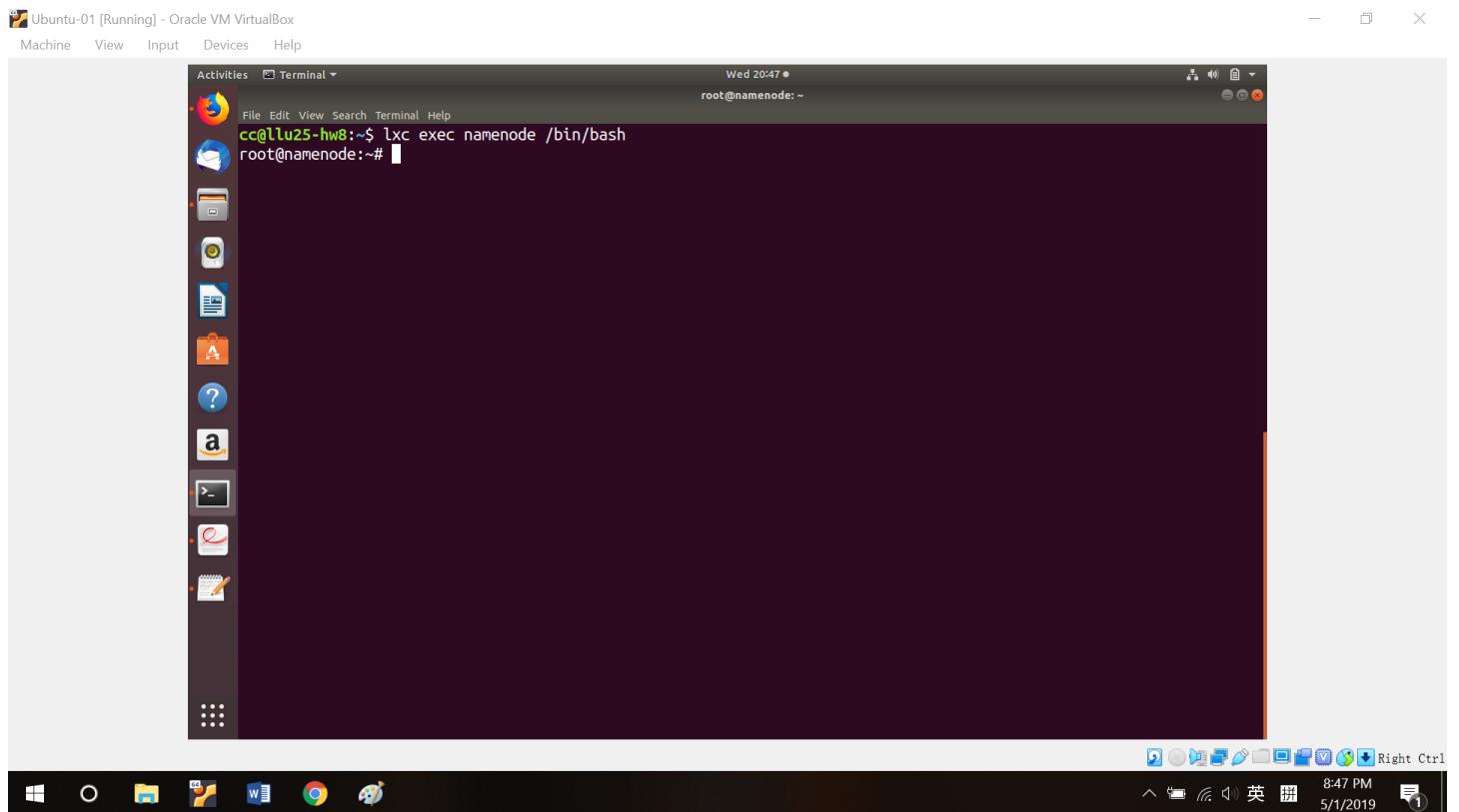
```
cc@llu25-hw8:~$ lxc launch ubuntu:16.04 namenode -c security.privileged=true -c raw.apparmor="mount fstype=rpc_piefs, mount fstype=nfsd, mount fstype=nfs,"  
Creating namenode  
Starting namenode  
cc@llu25-hw8:~$ lxc launch ubuntu:16.04 namenode -c security.privileged=true -c raw.apparmor="mount fstype=rpc_piefs, mount fstype=nfsd, mount fstype=nfs,"  
Creating namenode  
error: Failed container creation:  
- https://cloud-images.ubuntu.com/releases: Container 'namenode' already exists  
cc@llu25-hw8:~$ lxc launch ubuntu:16.04 datanode1 -c security.privileged=true -c raw.apparmor="mount fstype=rpc_piefs, mount fstype=nfsd, mount fstype=nfs,"  
Creating datanode1  
Starting datanode1  
cc@llu25-hw8:~$ lxc launch ubuntu:16.04 datanode2 -c security.privileged=true -c raw.apparmor="mount fstype=rpc_piefs, mount fstype=nfsd, mount fstype=nfs,"  
Creating datanode2  
Starting datanode2  
cc@llu25-hw8:~$ lxc launch ubuntu:16.04 datanode3 -c security.privileged=true -c raw.apparmor="mount fstype=rpc_piefs, mount fstype=nfsd, mount fstype=nfs,"  
Creating datanode3  
Starting datanode3  
cc@llu25-hw8:~$ lxc launch ubuntu:16.04 datanode4 -c security.privileged=true -c raw.apparmor="mount fstype=rpc_piefs, mount fstype=nfsd, mount fstype=nfs,"  
Creating datanode4  
Starting datanode4  
cc@llu25-hw8:~$
```

The desktop interface includes a dock with icons for various applications like a browser, file manager, and terminal, and a system tray at the bottom right showing the date and time.

(b) You can limit cpu number and memory space.

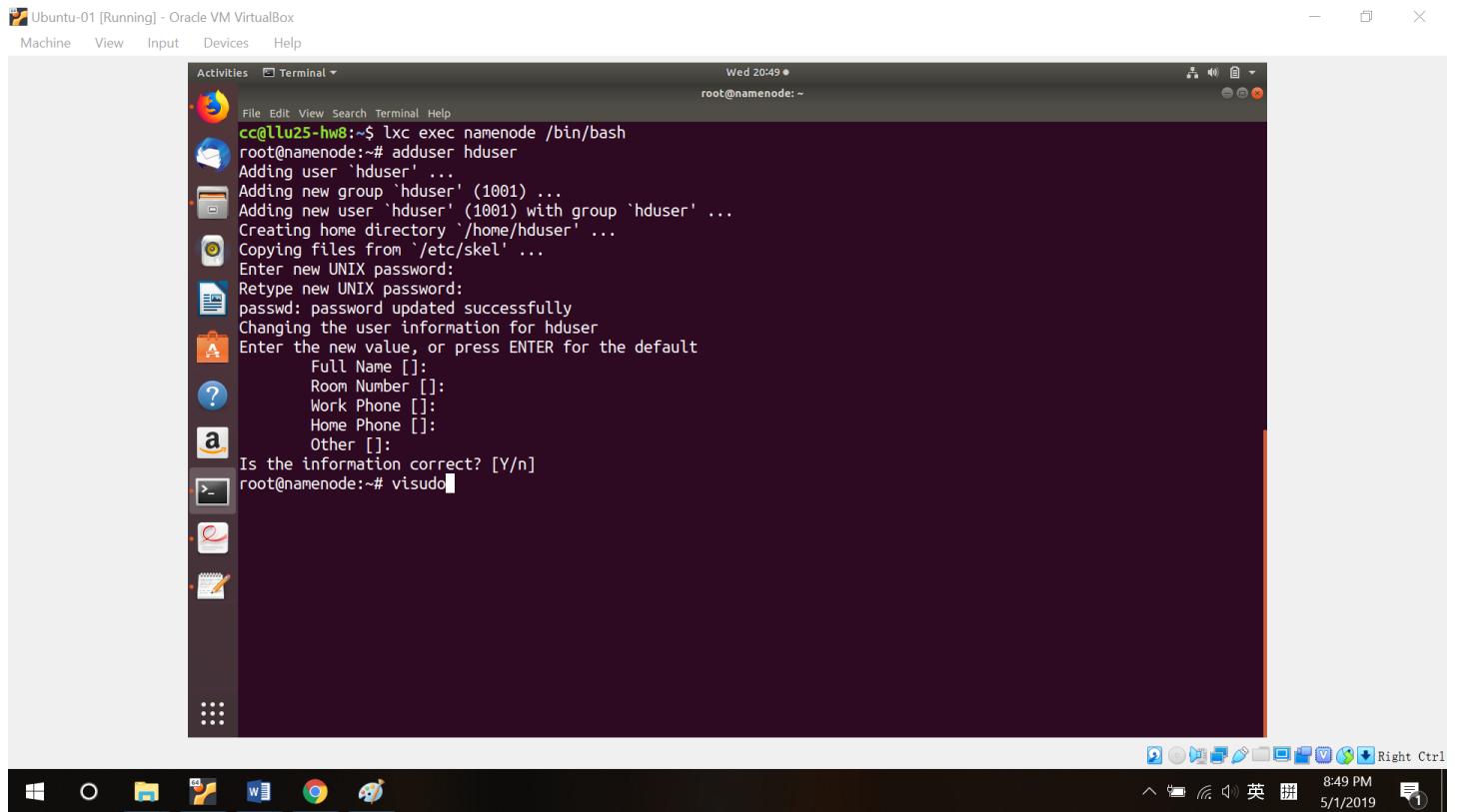


(c) Enter your virtual machine.

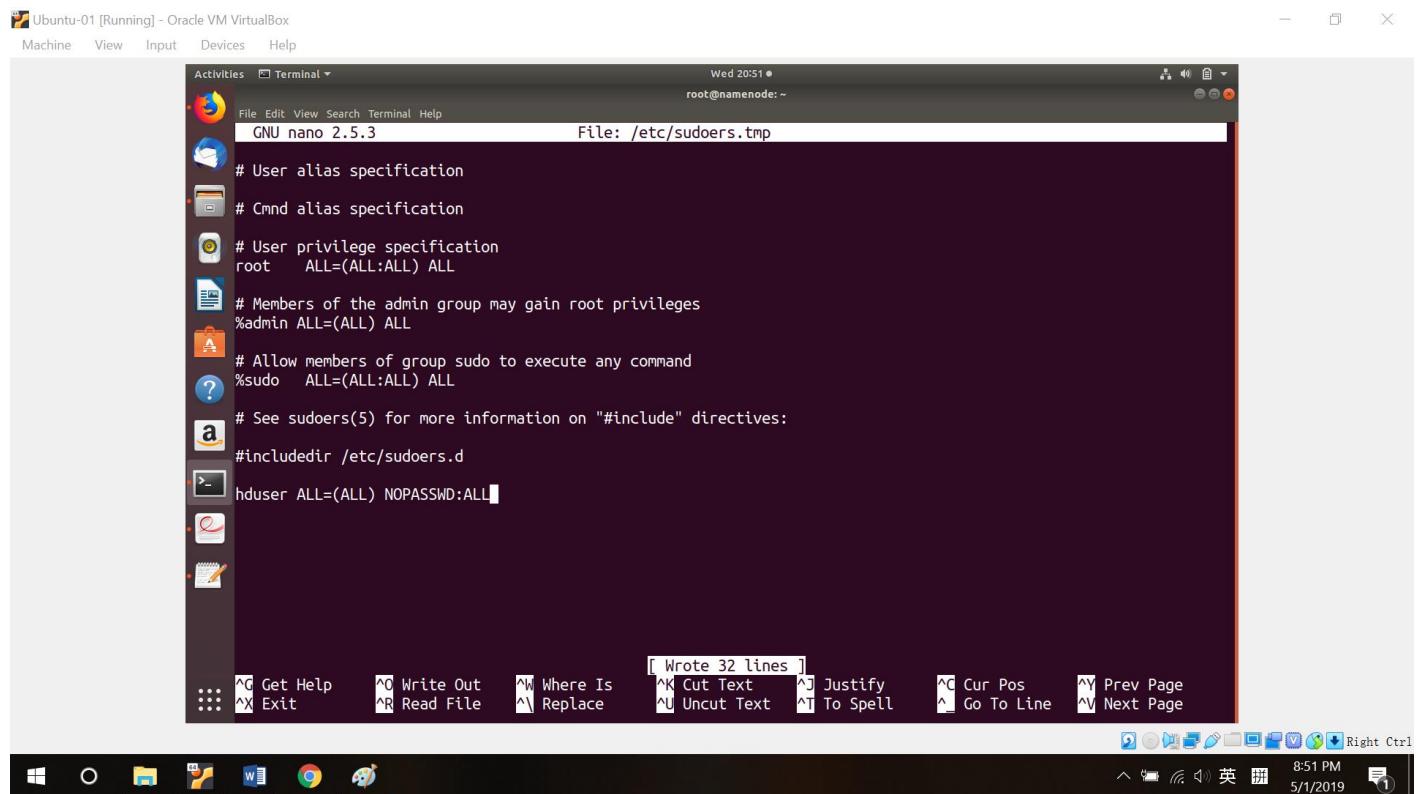


4: Add sudo user, install java and set up ssh.

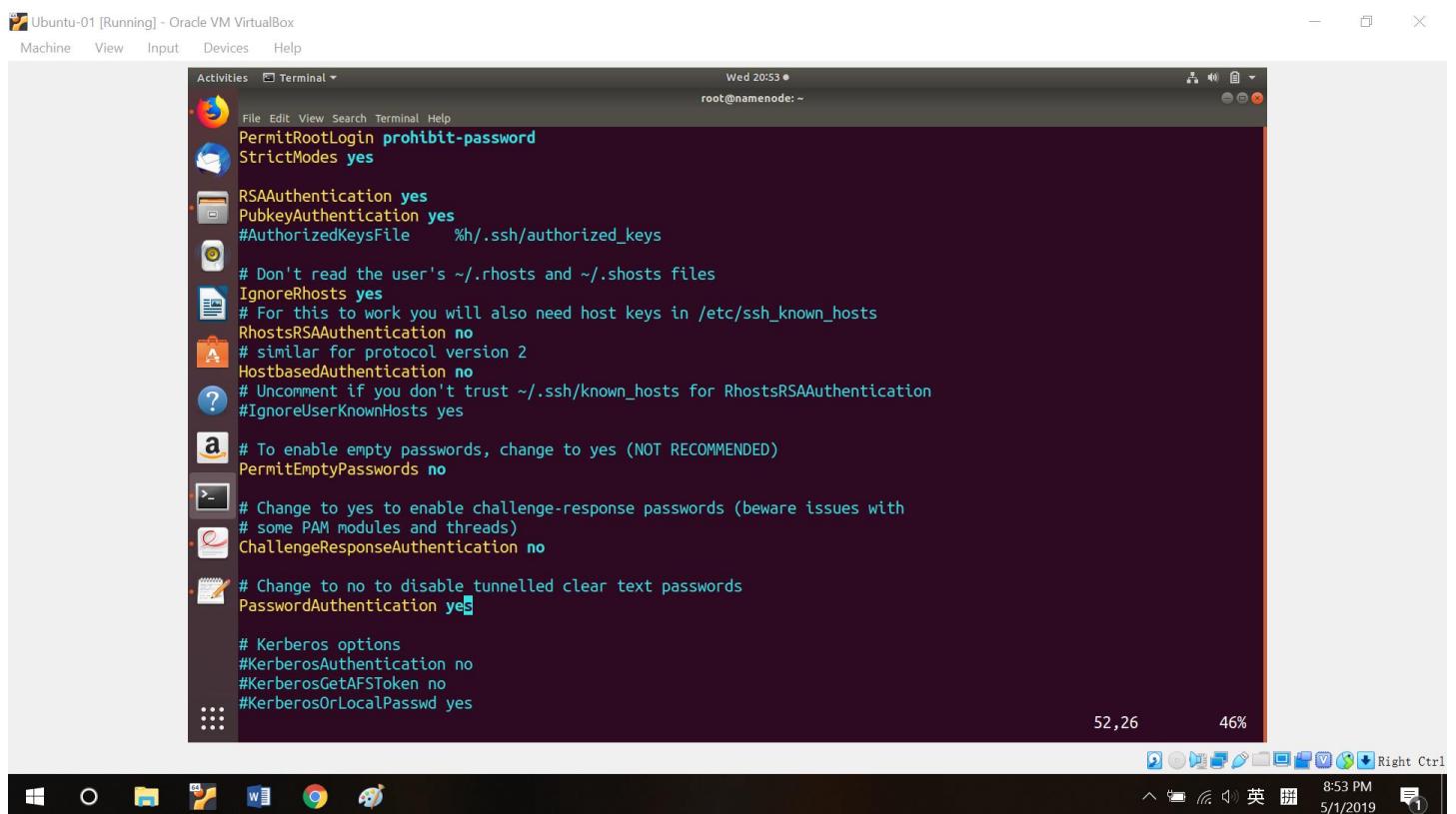
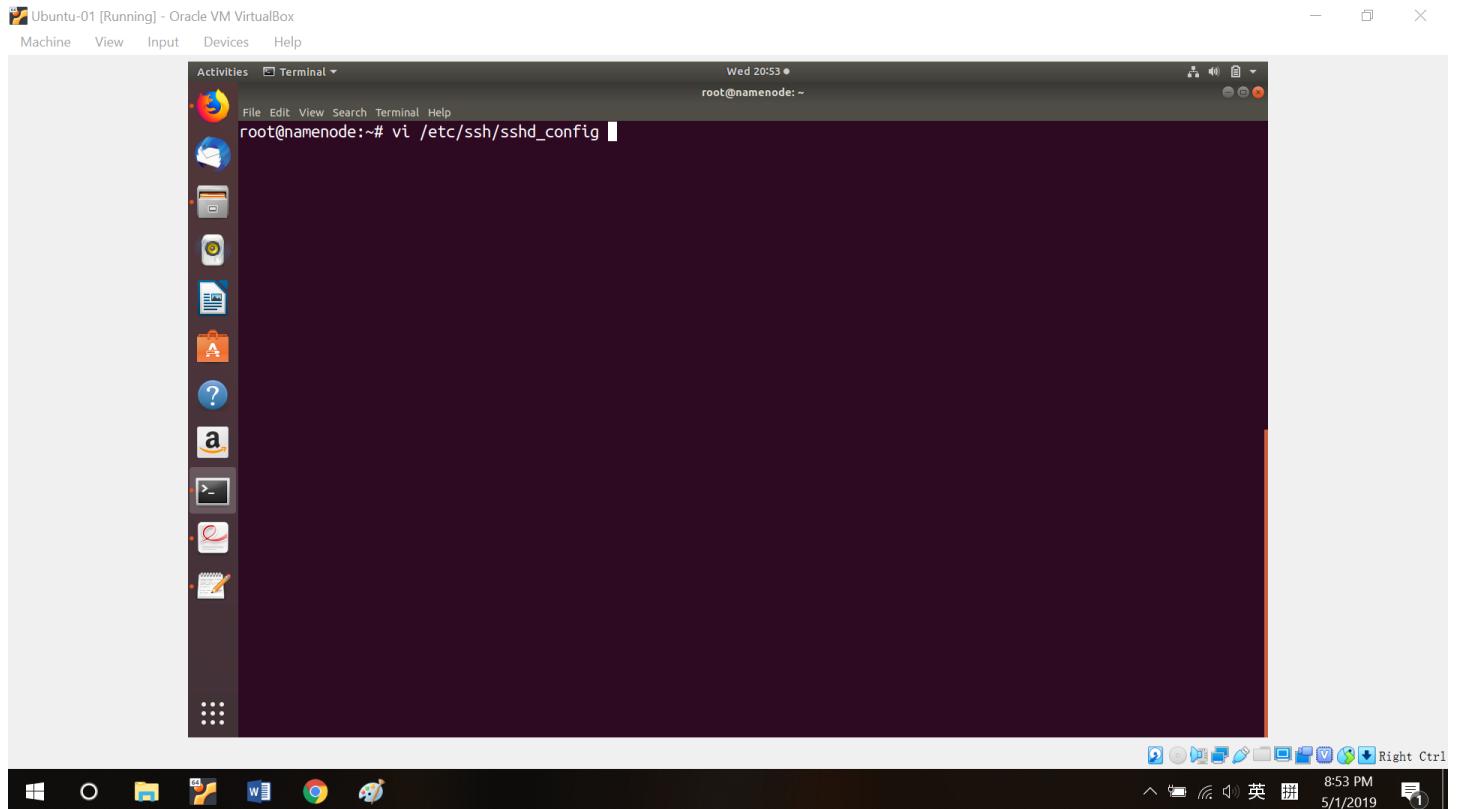
(a) I don't want to do everything with root permission. So I create sudo users.



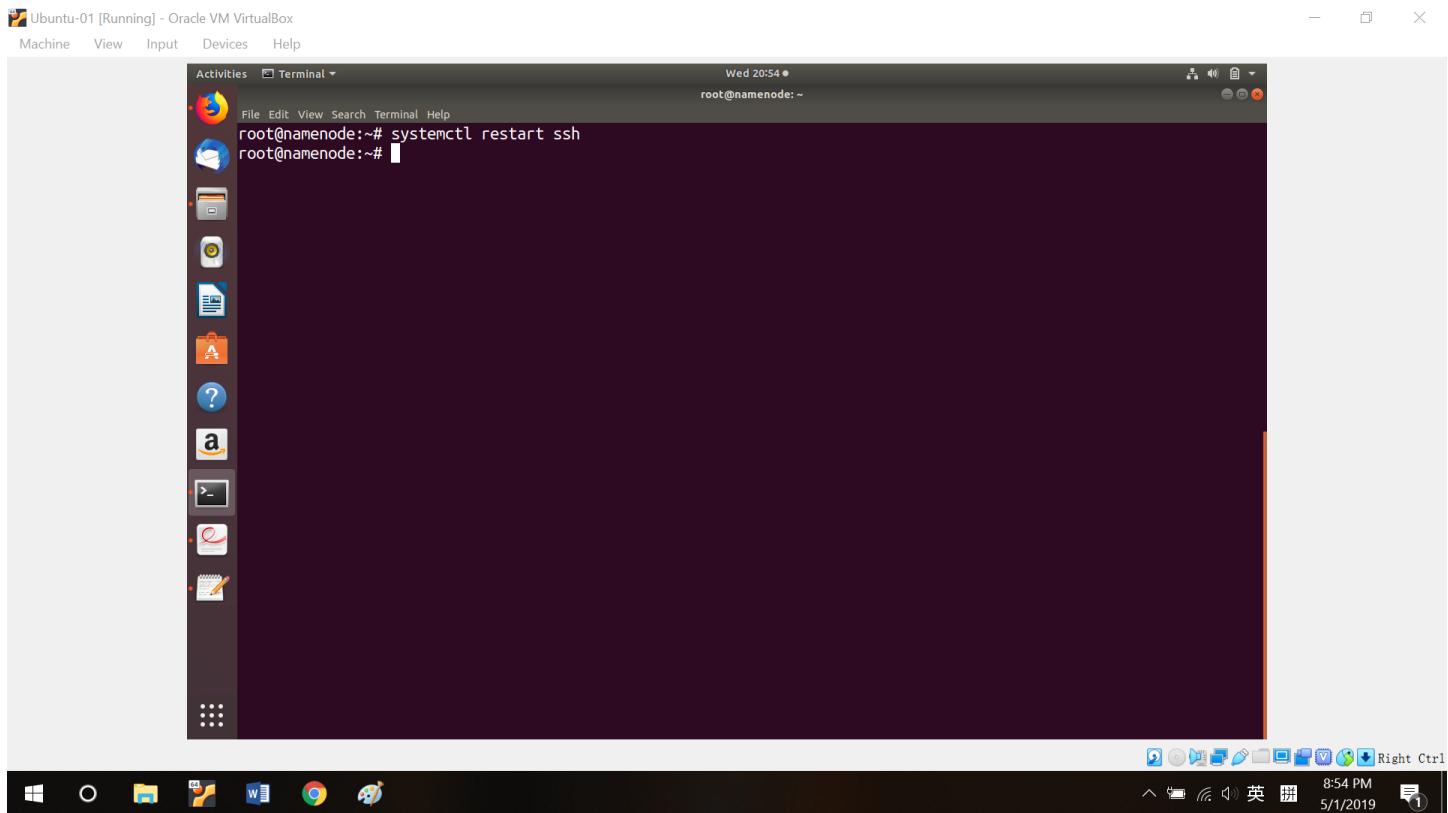
(b) Go into visudo to give the new created user sudo permission.



(c) lxc set password authorization to “no” by default. To use ssh, set it to “yes”.

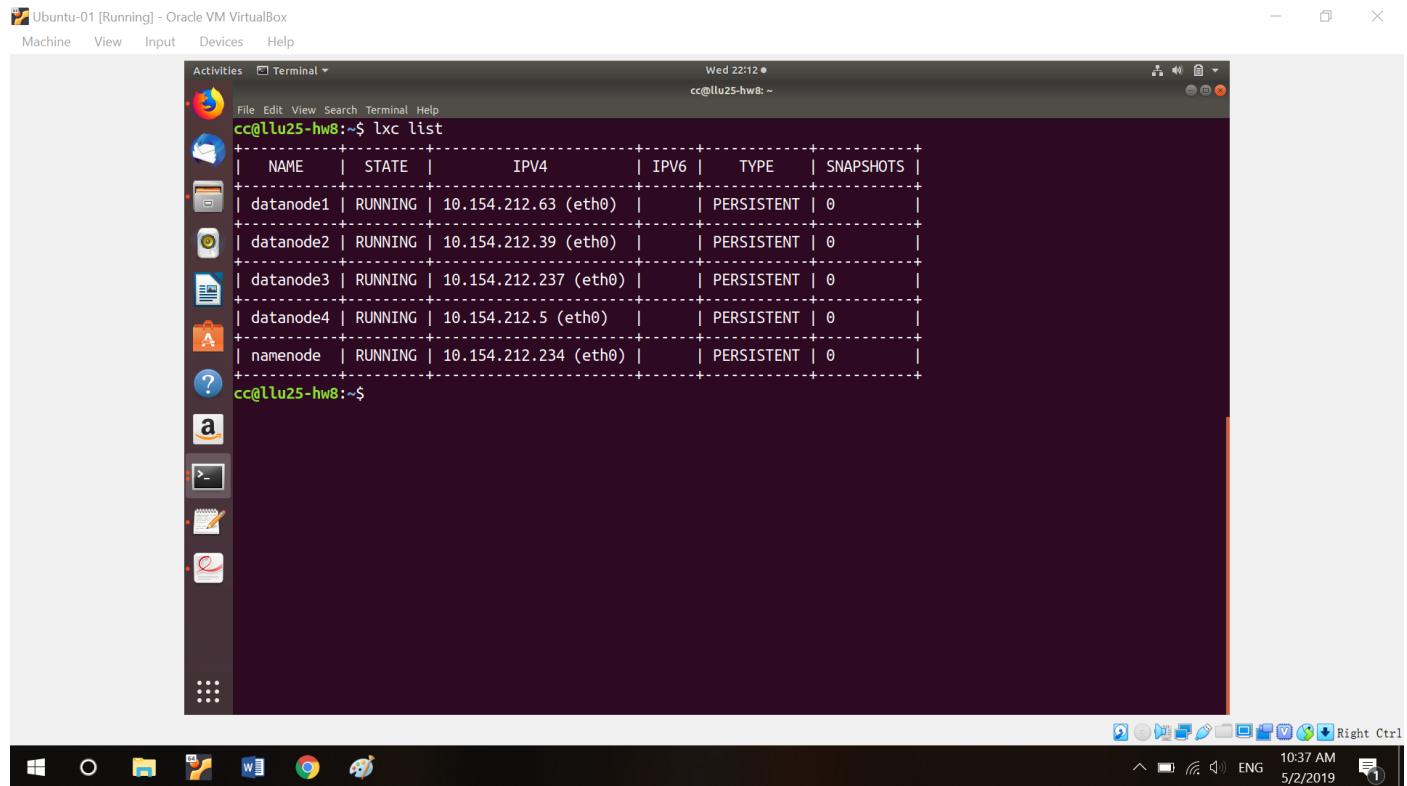


(d) Don't forget to restart ssh to active your setting.

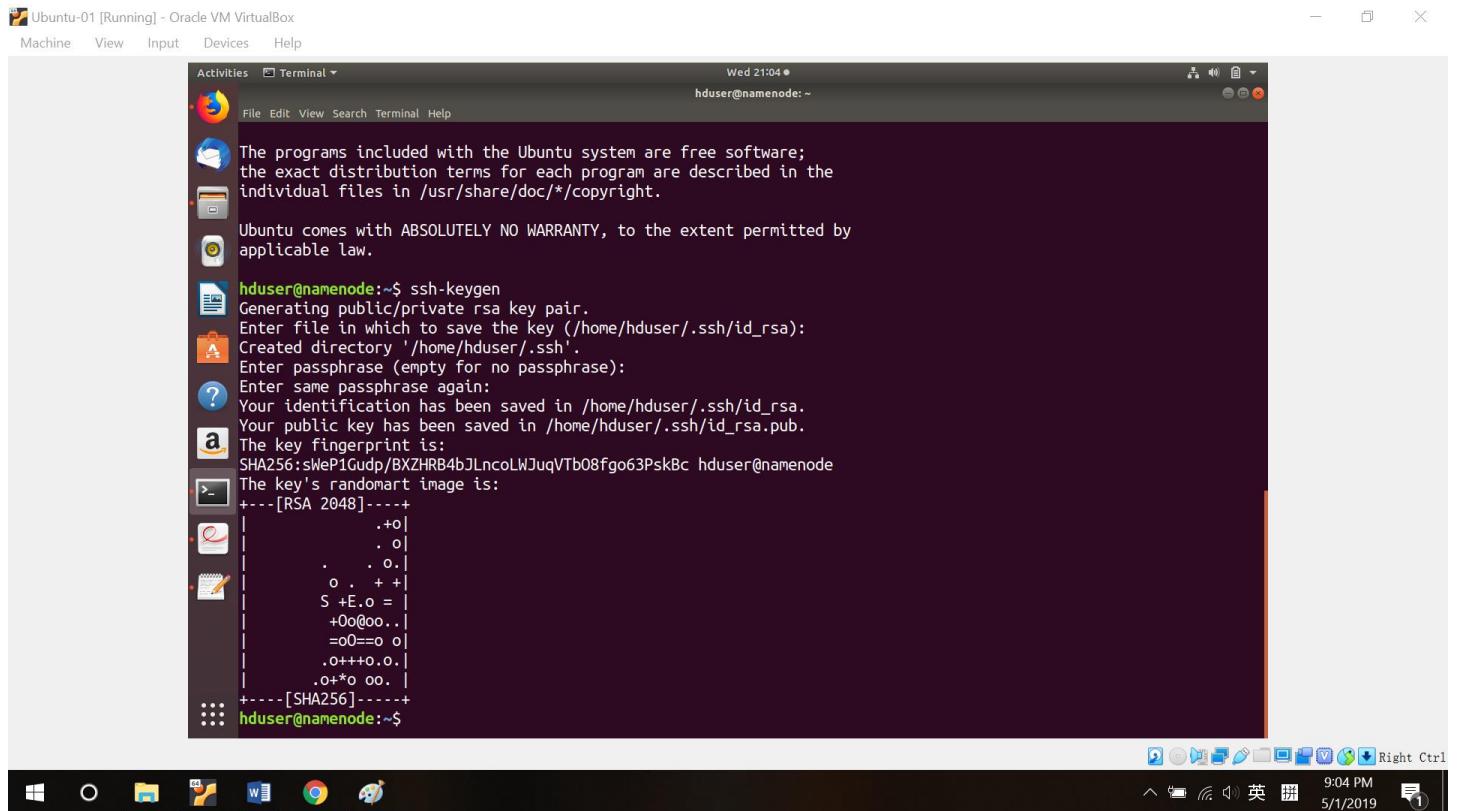


(e) Do step (a) to (d) on all nodes to allow ssh.

(f) Check all your nodes' IP addresses. Save them in a file. We will paste it into all nodes' host file. So we don't have to enter IP address every time.



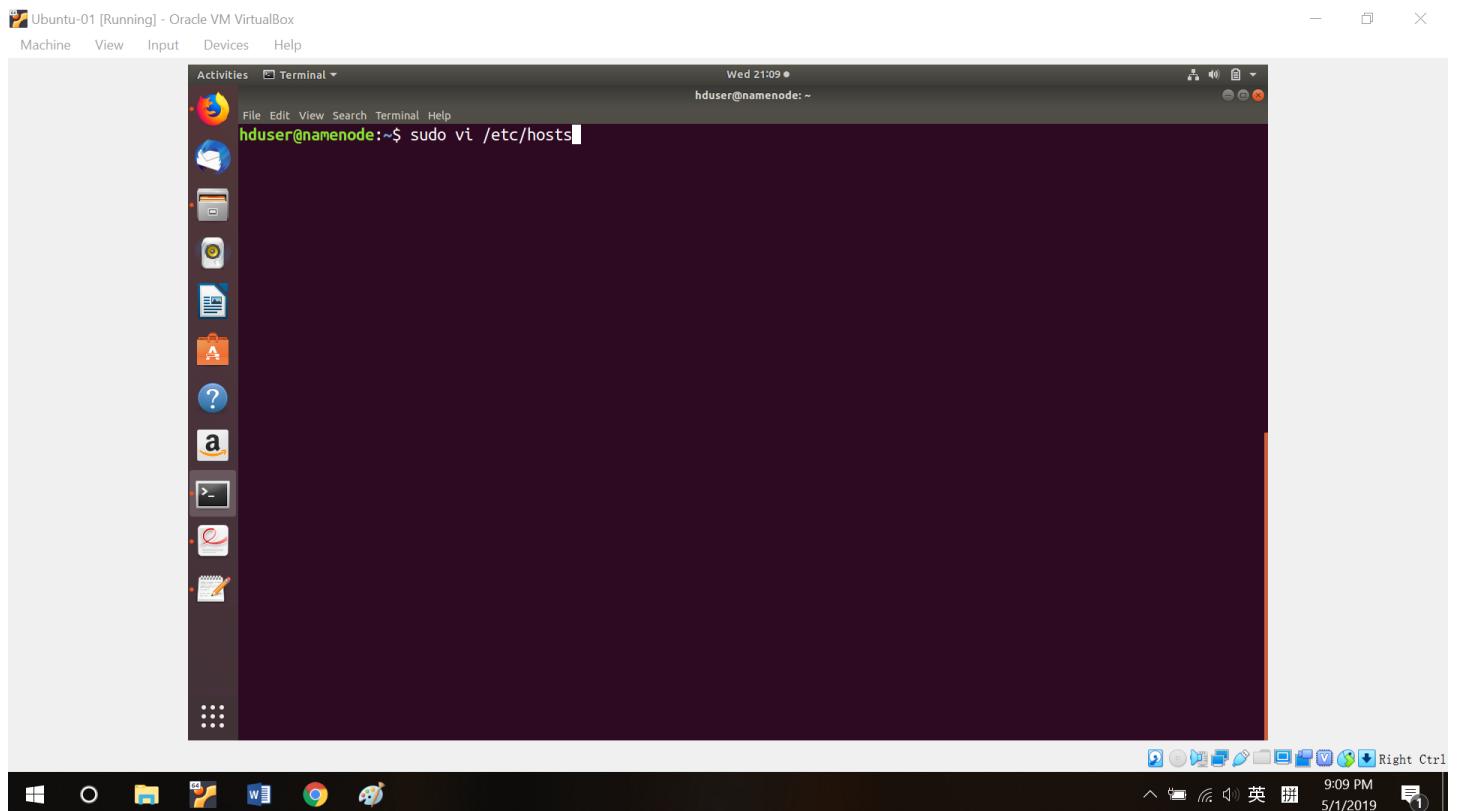
(g) Open your name node, update and upgrade your system (all nodes), generate a key pair (all nodes), overwrite host file (all nodes) and copy your public key to other nodes (name node must copy its key to all nodes which includes itself; data node only needs to send their public key to name node).



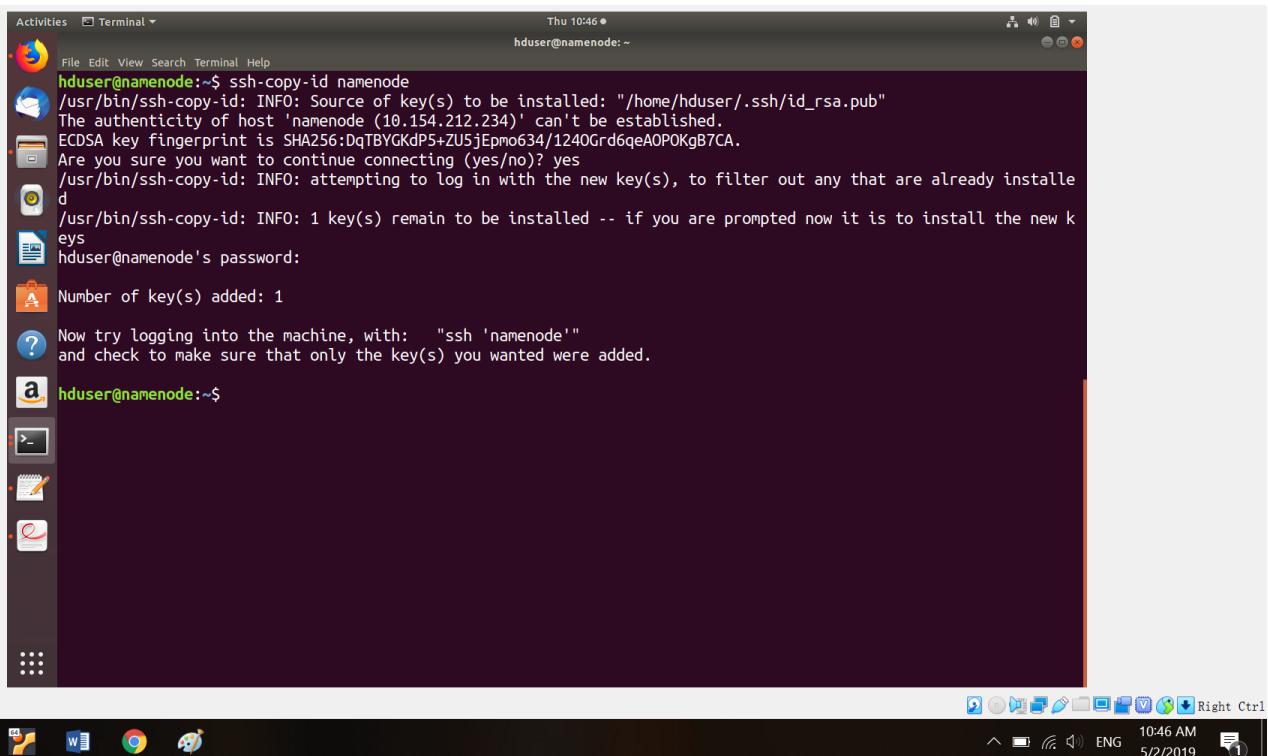
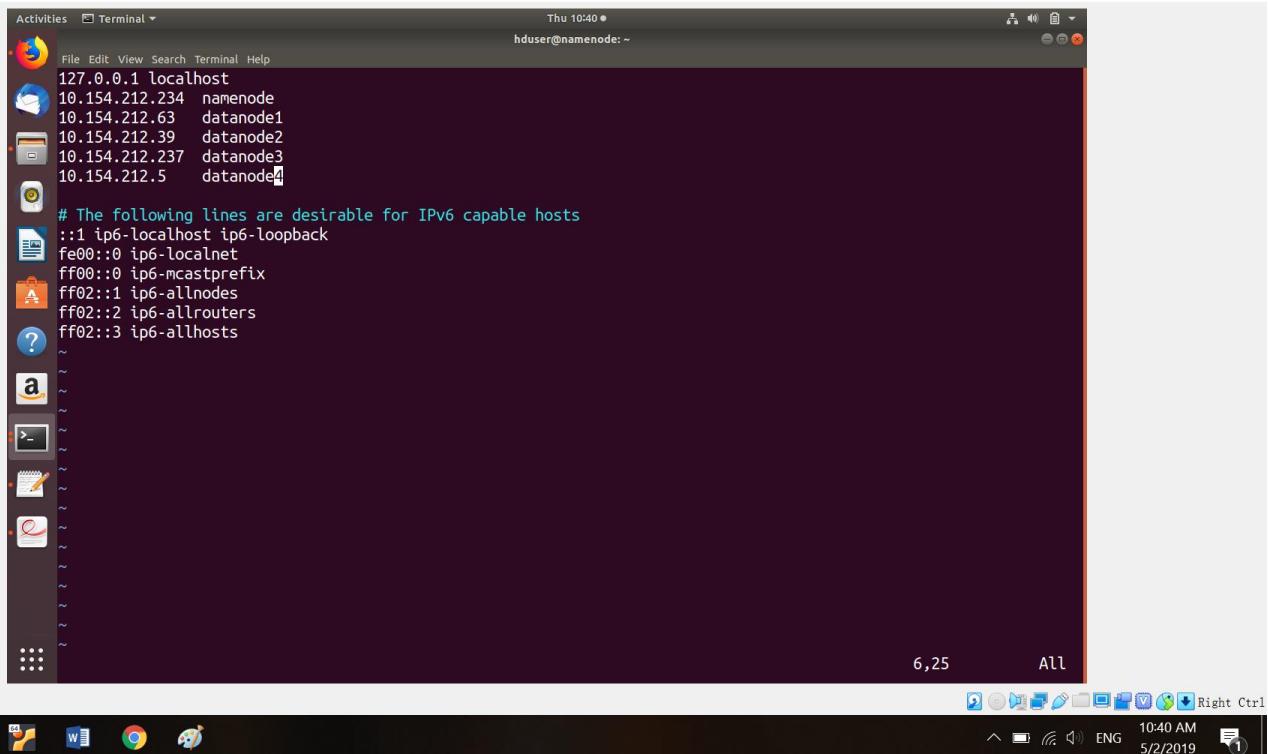
```
The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*copyright.

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

hduser@namenode:~$ ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hduser/.ssh/id_rsa):
Created directory '/home/hduser/.ssh'.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
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:sWeP1Gudp/BXZHR84bJLnc0LWJuqVTb08fg063PskBc hduser@namenode
The key's randomart image is:
+---[RSA 2048]----+
|          .+o|
|          . o|
|          . o.|
|          o . + +|
|          S +E.o =|
|          +0o@oo..|
|          =oO==o o|
|          .o+++o.o.|
|          .o*o oo.|
+---[SHA256]----+
hduser@namenode:~$
```



```
hduser@namenode:~$ sudo vi /etc/hosts
```



Ubuntu-01 [Running] - Oracle VM VirtualBox

Machine View Input Devices Help

Activities Terminal Thu 10:50 ●
hduser@datanode4: ~

```
The key fingerprint is:  
SHA256:xMUEzipxLE1GM+pgCAdH3TasF8b1/TJZQ/65AmvvoY8 hduser@datanode4  
The key's randomart image is:  
+---[RSA 2048]---+  
|o0+. =*..o+ . |  
|.+ ..*o* o.. o |  
. o ++o= . . + |  
. o..o . + o.|  
o.. S . + ...|  
. o+ . |  
o.... |  
.o.. |  
Eoo |  
+---[SHA256]---+  
hduser@datanode4:~$ ssh-copy-id namenode  
/usr/bin/ssh-copy-id: INFO: Source of key(s) to be installed: "/home/hduser/.ssh/id_rsa.pub"  
The authenticity of host 'namenode (10.154.212.234)' can't be established.  
ECDSA key fingerprint is SHA256:DqTBYGKdP5+ZU5jEpno634/1240Grd6qeAOPOKgB7CA.  
Are you sure you want to continue connecting (yes/no)? yes  
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s), to filter out any that are already installed  
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you are prompted now it is to install the new keys  
hduser@namenode's password:  
Number of key(s) added: 1  
  
Now try logging into the machine, with: "ssh 'namenode'"  
and check to make sure that only the key(s) you wanted were added.  
hduser@datanode4:~$
```

Right Ctrl

Windows Start File Explorer Microsoft Word Microsoft Edge Task View

10:50 AM 5/2/2019 ENG

(f) Install java and set up JAVA_HOME on all nodes.

Ubuntu-01 [Running] - Oracle VM VirtualBox

Machine View Input Devices Help

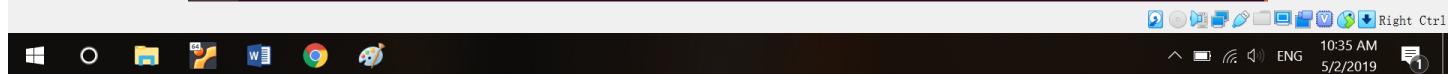
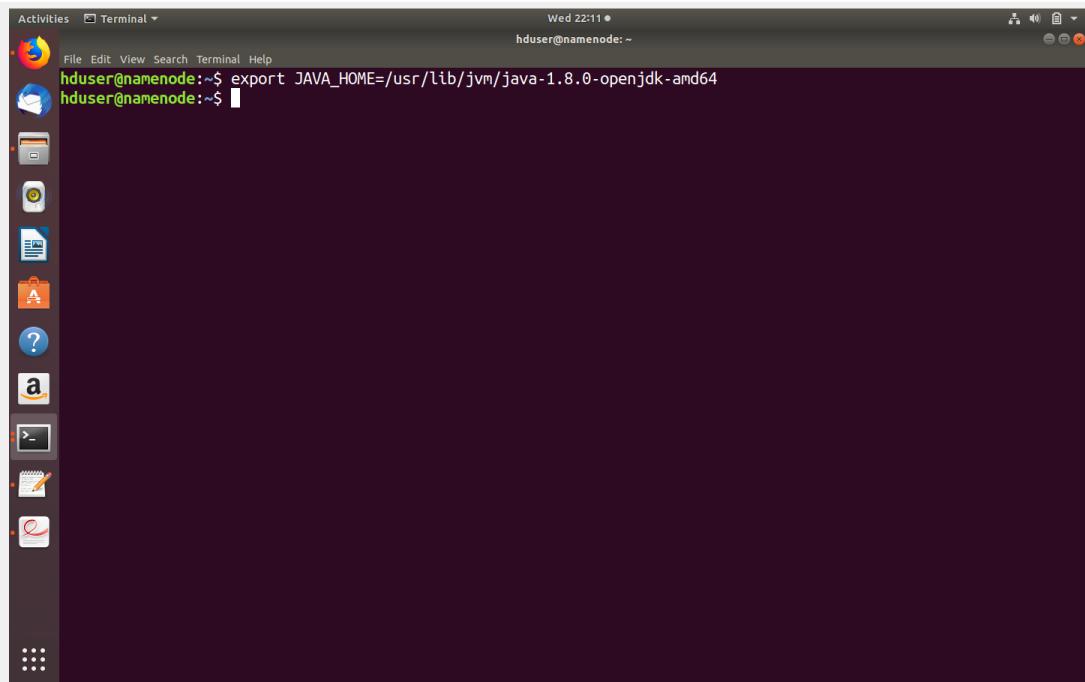
Activities Terminal Wed 22:08 ●
hduser@namenode: ~

```
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
The following additional packages will be installed:  
libpulse0 libsm-dev libx11-dev libx11-doc libxau-dev libxcb1-dev libxdmcp-dev libxt-dev libxt6  
openjdk-8-jdk-headless openjdk-8-jre x11proto-input-dev x11proto-kb-dev xtrans-dev  
Suggested packages:  
pulseaudio libsm-doc libxcb-doc libx11-doc openjdk-8-demo openjdk-8-source visualvm icedtea-8-plugin  
The following NEW packages will be installed:  
libpulse0 libsm-dev libx11-dev libx11-doc libxau-dev libxcb1-dev libxdmcp-dev libxt-dev libxt6 openjdk-8-jdk  
openjdk-8-jdk-headless openjdk-8-jre x11proto-input-dev x11proto-kb-dev xtrans-dev  
0 upgraded, 15 newly installed, 0 to remove and 0 not upgraded.  
Need to get 0 B/13.2 MB of archives.  
After this operation, 59.2 MB of additional disk space will be used.  
Do you want to continue? [Y/n] y  
Selecting previously unselected package libpulse0:amd64.  
(Reading database ... 27293 files and directories currently installed.)  
Preparing to unpack .../libpulse0_1%3a8.0-0ubuntu3.10_amd64.deb ...  
Unpacking libpulse0:amd64 (1:8.0-0ubuntu3.10) ...  
Selecting previously unselected package libsm-dev:amd64.  
Preparing to unpack .../libsm-dev_2%3a1.2.2-1_amd64.deb ...  
Unpacking libsm-dev:amd64 (2:1.2.2-1) ...  
Selecting previously unselected package libxau-dev:amd64.  
Preparing to unpack .../libxau-dev_1%3a1.0.8-1_amd64.deb ...  
Unpacking libxau-dev:amd64 (1:1.0.8-1) ...  
Selecting previously unselected package libxdmcp-dev:amd64.  
Preparing to unpack .../libxdmcp-dev_1%3a1.1.2-1.1_amd64.deb ...  
Unpacking libxdmcp-dev:amd64 (1:1.1.2-1.1) ...  
Selecting previously unselected package x11proto-input-dev.  
Preparing to unpack .../x11proto-input-dev_2.3.1-1_all.deb ...  
Unpacking x11proto-input-dev (2.3.1-1)
```

Right Ctrl

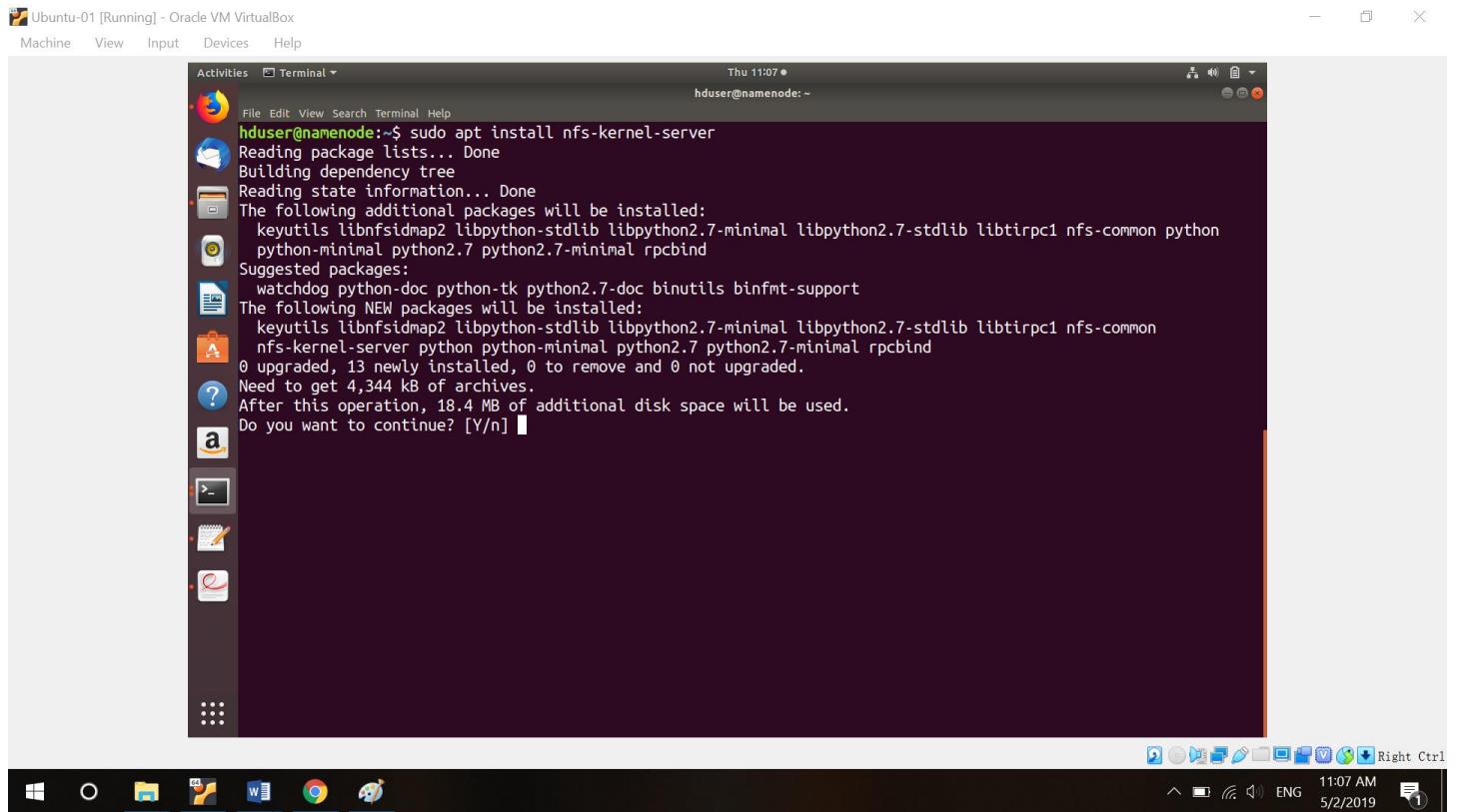
Windows Start File Explorer Microsoft Word Microsoft Edge Task View

10:33 AM 5/2/2019 ENG



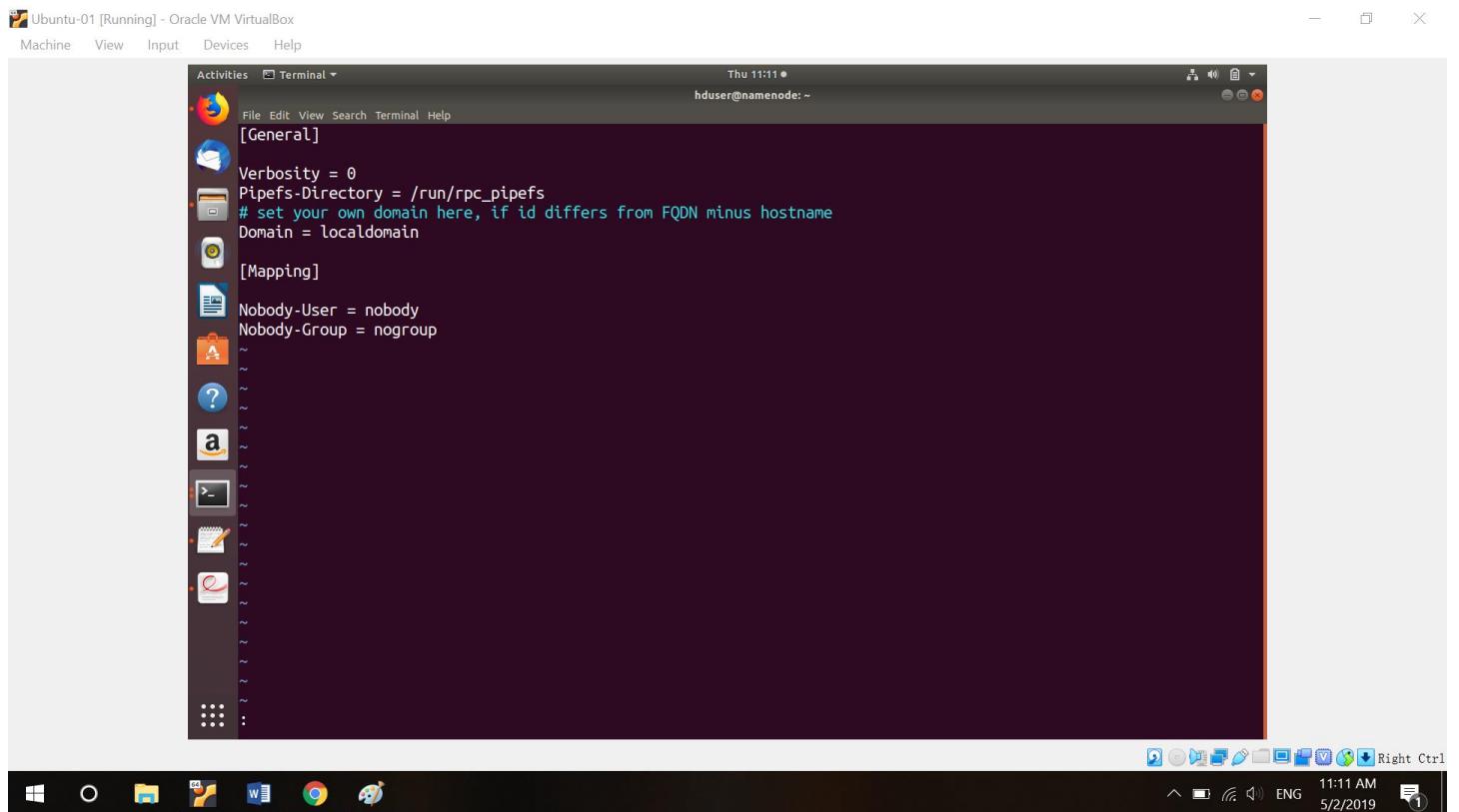
5. NFS

(a) Install nfs server on your name node. And open idmapd.conf and uncomment line 6.



A screenshot of an Ubuntu desktop environment running in Oracle VM VirtualBox. The terminal window shows the command `sudo apt install nfs-kernel-server` being run, and the output of the package manager. The desktop interface includes a dock with various icons and a system tray at the bottom right.

```
hduser@namenode:~$ sudo apt install nfs-kernel-server
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  keyutils libnfsidmap2 libpython-stplib libpython2.7-minimal libpython2.7-stplib libtirpc1 nfs-common python
    python-minimal python2.7 python2.7-minimal rpcbind
  Suggested packages:
    watchdog python-doc python-tk python2.7-doc binutils bincfmt-support
  The following NEW packages will be installed:
  keyutils libnfsidmap2 libpython-stplib libpython2.7-minimal libpython2.7-stplib libtirpc1 nfs-common
    nfs-kernel-server python python-minimal python2.7 python2.7-minimal rpcbind
0 upgraded, 13 newly installed, 0 to remove and 0 not upgraded.
Need to get 4,344 kB of archives.
After this operation, 18.4 MB of additional disk space will be used.
Do you want to continue? [Y/n] ■
```

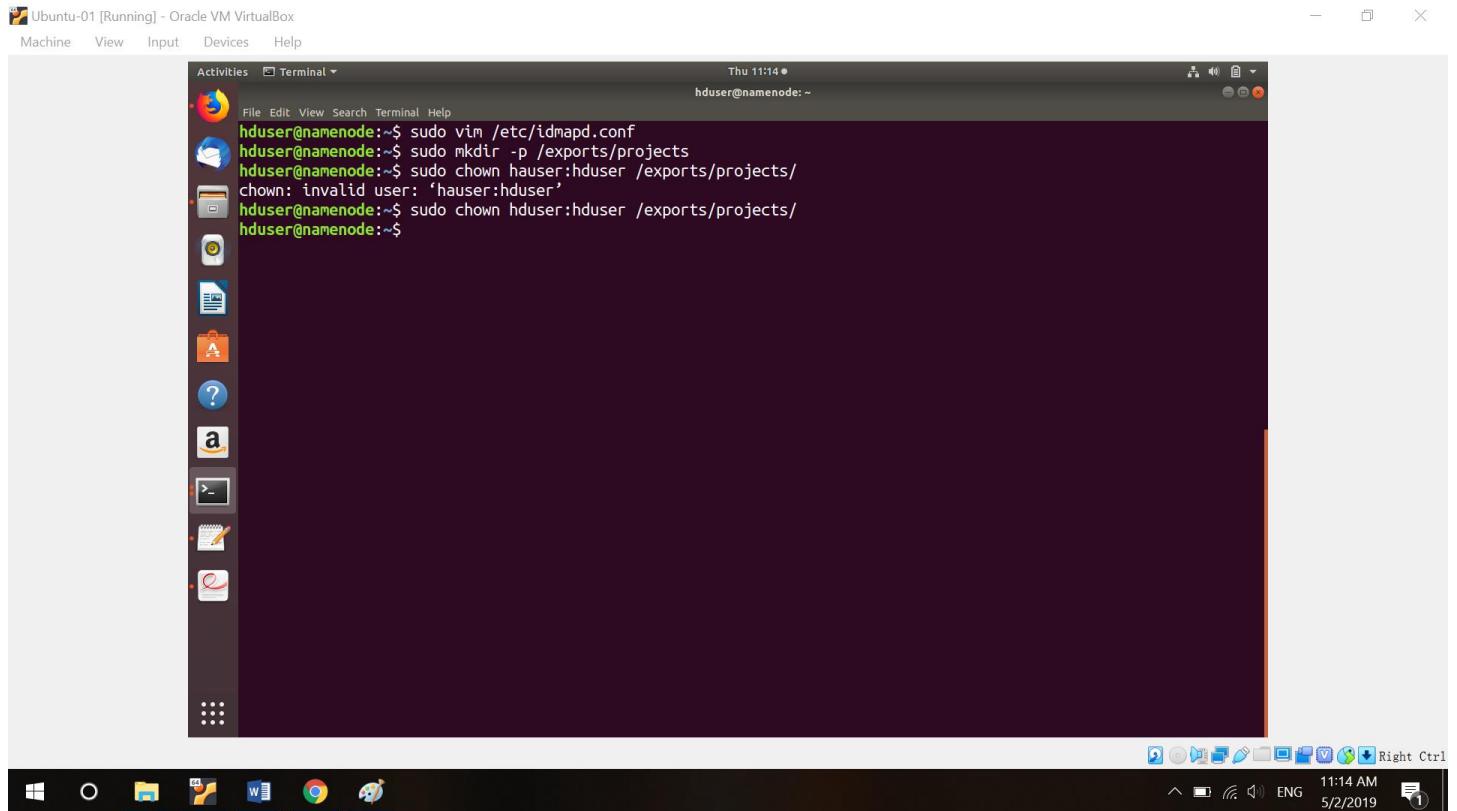


A screenshot of an Ubuntu desktop environment running in Oracle VM VirtualBox. The terminal window shows the configuration file `/etc/idmapd.conf` with several lines commented out. The desktop interface includes a dock with various icons and a system tray at the bottom right.

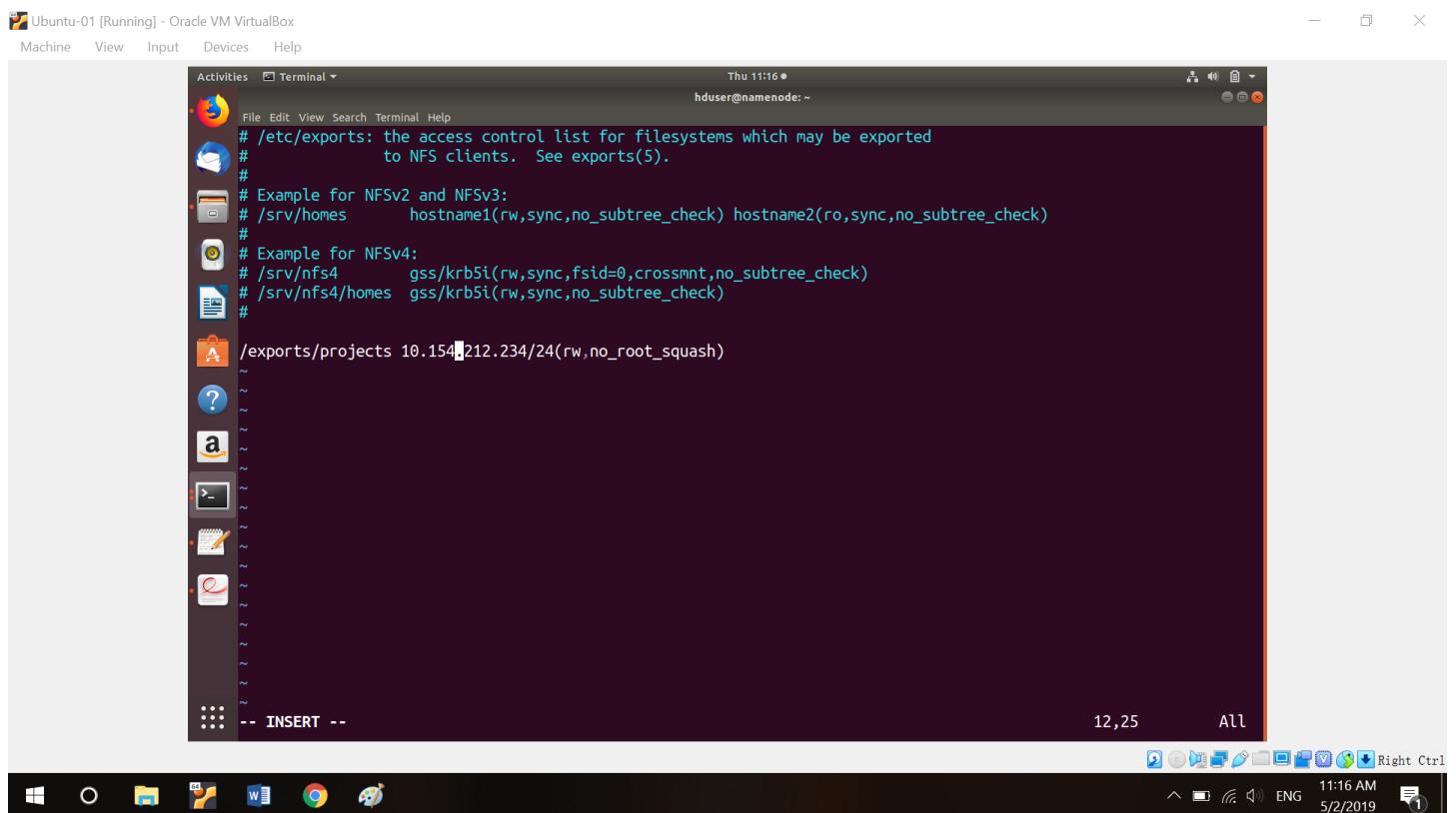
```
[General]
Verbosity = 0
Pipesfs-Directory = /run/rpc_pipefs
# set your own domain here, if id differs from FQDN minus hostname
Domain = localdomain

[Mapping]
Nobody-User = nobody
Nobody-Group = nogroup
```

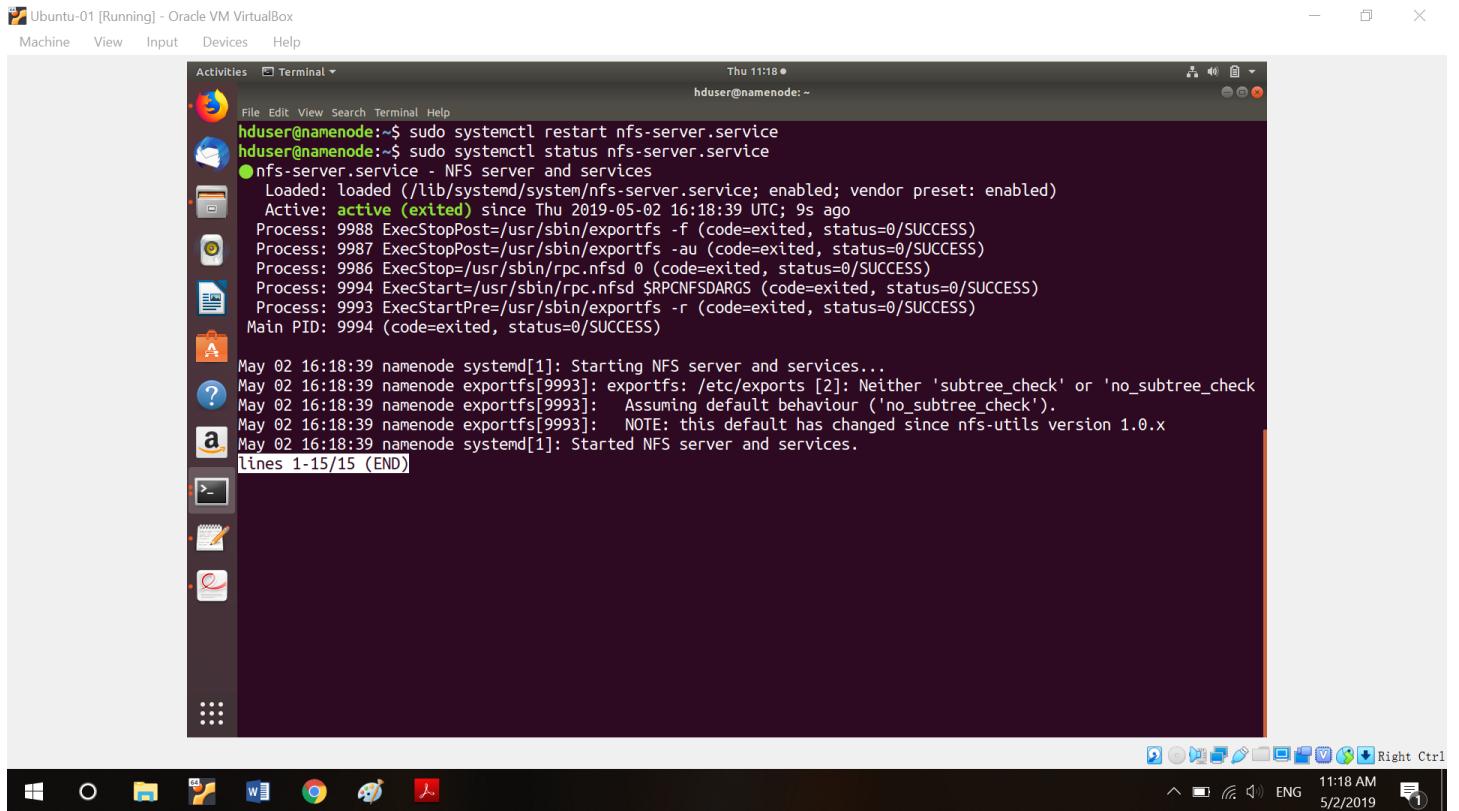
(b) Create a directory for nfs and give permission to current user.



(c) Open “sudo vim /etc/exports” and change like this. IP address is your name node’s IP address. (24 means maximize network permission). This command allows 10.154.212.234/24 to mount the file.



(d) Restart your nfs server and see its current status. It should look like this.



Ubuntu-01 [Running] - Oracle VM VirtualBox

Machine View Input Devices Help

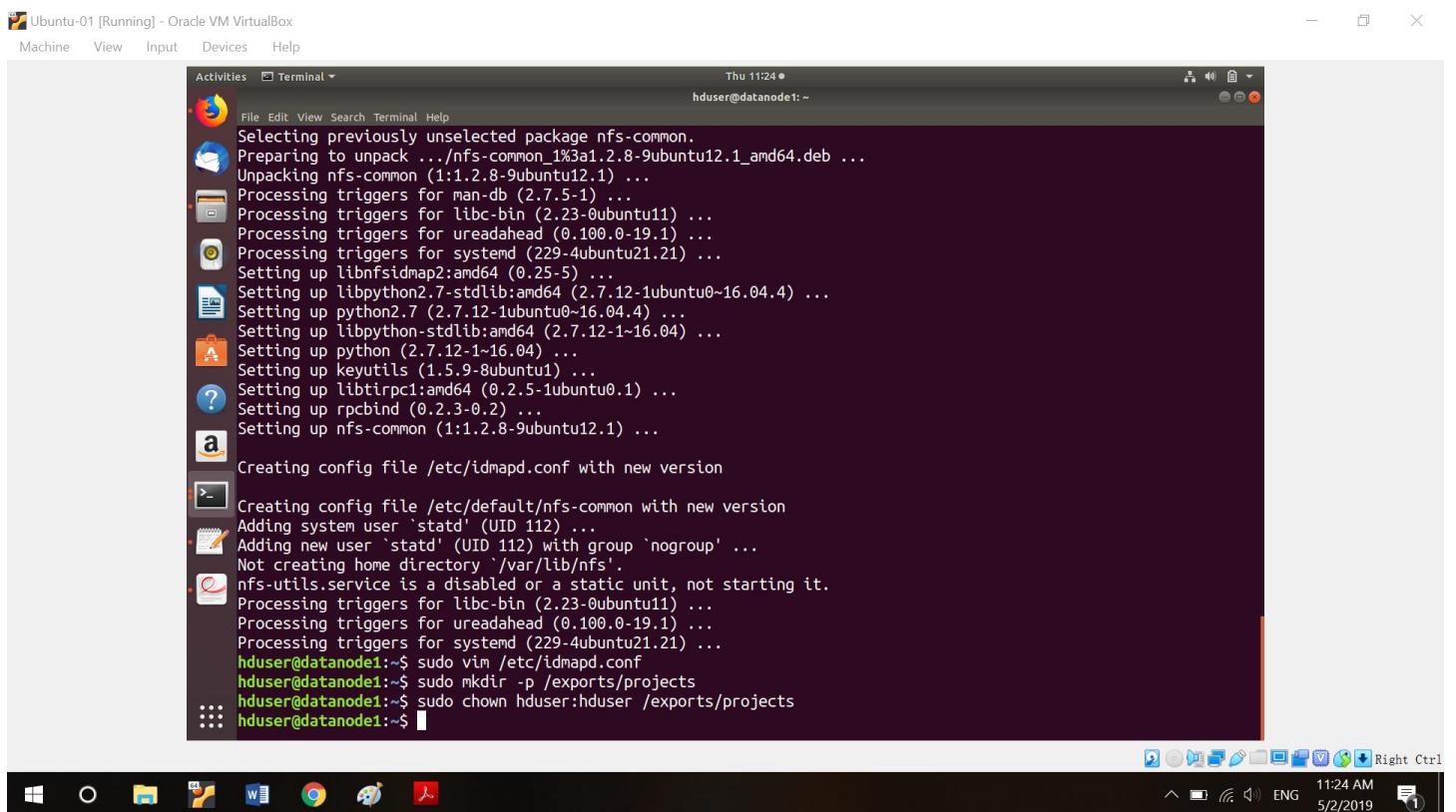
Activities Terminal Thu 11:18 ● hduser@namenode: ~

```
hduser@namenode:~$ sudo systemctl restart nfs-server.service
hduser@namenode:~$ sudo systemctl status nfs-server.service
● nfs-server.service - NFS server and services
   Loaded: loaded (/lib/systemd/system/nfs-server.service; enabled; vendor preset: enabled)
   Active: active (exited) since Thu 2019-05-02 16:18:39 UTC; 9s ago
     Process: 9988 ExecStopPost=/usr/sbin/exportfs -f (code=exited, status=0/SUCCESS)
     Process: 9987 ExecStopPost=/usr/sbin/exportfs -au (code=exited, status=0/SUCCESS)
     Process: 9986 ExecStop=/usr/sbin/rpc.nfsd 0 (code=exited, status=0/SUCCESS)
     Process: 9994 ExecStart=/usr/sbin/rpc.nfsd $RPCNFSDARGS (code=exited, status=0/SUCCESS)
     Process: 9993 ExecStartPre=/usr/sbin/exportfs -r (code=exited, status=0/SUCCESS)
   Main PID: 9994 (code=exited, status=0/SUCCESS)

May 02 16:18:39 namenode systemd[1]: Starting NFS server and services...
May 02 16:18:39 namenode exportfs[9993]: exportfs: /etc/exports [2]: Neither 'subtree_check' or 'no_subtree_check'
May 02 16:18:39 namenode exportfs[9993]: Assuming default behaviour ('no_subtree_check').
May 02 16:18:39 namenode exportfs[9993]: NOTE: this default has changed since nfs-utils version 1.0.x
May 02 16:18:39 namenode systemd[1]: Started NFS server and services.
lines 1-15/15 (END)
```

Windows Start button, Dash icon, File Explorer, Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Microsoft Edge, File Explorer, Taskbar icons, System tray: ENG 11:18 AM 5/2/2019

(c) Go to your data nodes, download nfs client, change idmapd.conf, create a new directory (/exports/projects, name must totally equal to name node directory), and give current user permission.



Ubuntu-01 [Running] - Oracle VM VirtualBox

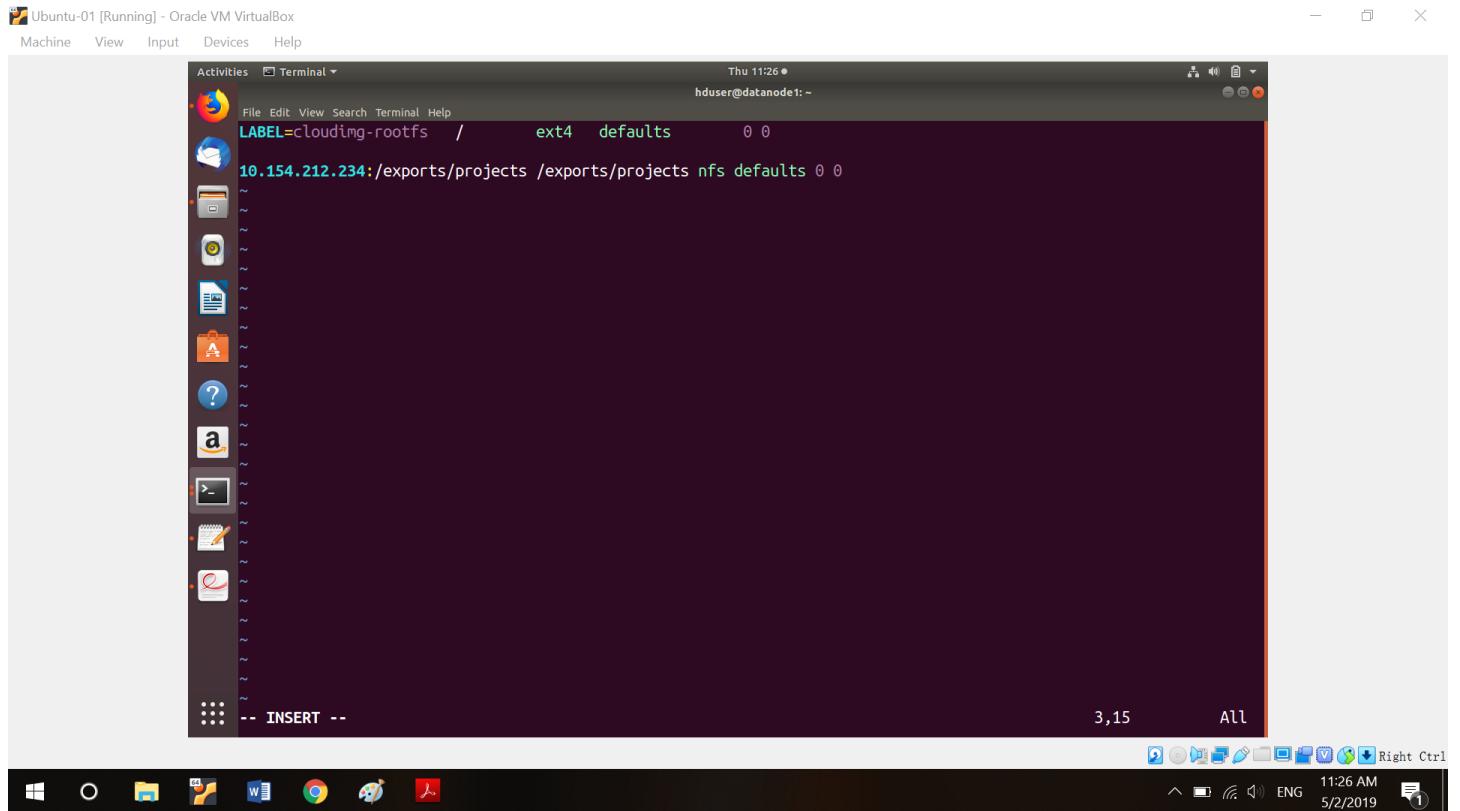
Machine View Input Devices Help

Activities Terminal Thu 11:24 ● hduser@datanode1: ~

```
Selecting previously unselected package nfs-common.
Preparing to unpack .../nfs-common_1%3a1.2.8-9ubuntu12.1_amd64.deb ...
Unpacking nfs-common (1:1.2.8-9ubuntu12.1) ...
Processing triggers for man-db (2.7.5-1) ...
Processing triggers for libc-bin (2.23-0ubuntu11) ...
Processing triggers for ureadahead (0.100.0-19.1) ...
Processing triggers for systemd (229-4ubuntu21.21) ...
Setting up libnfsidmap2:amd64 (0.25-5) ...
Setting up libpython2.7-stdlib:amd64 (2.7.12-1ubuntu0~16.04.4) ...
Setting up libpython-stdlib:amd64 (2.7.12-1-16.04) ...
Setting up python (2.7.12-1-16.04) ...
Setting up keyutils (1.5.9-8ubuntu1) ...
Setting up libtirpc1:amd64 (0.2.5-1ubuntu0.1) ...
Setting up rpcbind (0.2.3-0.2) ...
Setting up nfs-common (1:1.2.8-9ubuntu12.1) ...
Creating config file /etc/idmapd.conf with new version
Creating config file /etc/default/nfs-common with new version
Adding system user `statd' (UID 112) ...
Adding new user `statd' (UID 112) with group `nogroup' ...
Not creating home directory `/var/lib/nfs'.
nfs-utils.service is a disabled or a static unit, not starting it.
Processing triggers for libc-bin (2.23-0ubuntu11) ...
Processing triggers for ureadahead (0.100.0-19.1) ...
Processing triggers for systemd (229-4ubuntu21.21) ...
hduser@datanode1:~$ sudo vim /etc/idmapd.conf
hduser@datanode1:~$ sudo mkdir -p /exports/projects
hduser@datanode1:~$ sudo chown hduser:hduser /exports/projects
hduser@datanode1:~$
```

Windows Start button, Dash icon, File Explorer, Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Microsoft Edge, File Explorer, Taskbar icons, System tray: ENG 11:24 AM 5/2/2019

(d) Open “sudo vim /etc/fstab” and modify the like this.

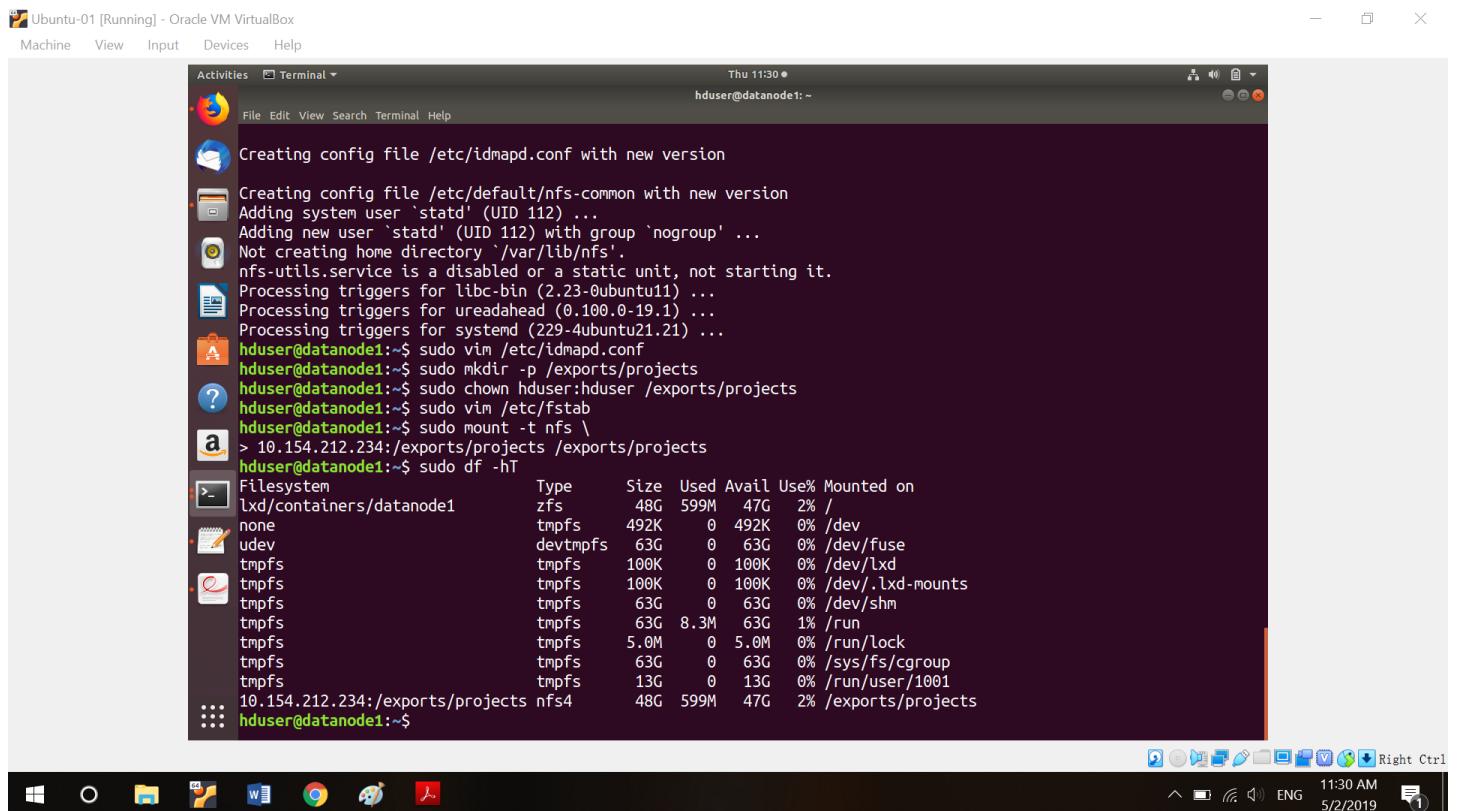


The screenshot shows a terminal window titled "Terminal" with the command "hduser@datanode1: ~". The content of the terminal is as follows:

```
File Edit View Search Terminal Help
LABEL=cloudimg-rootfs  /      ext4  defaults    0 0
10.154.212.234:/exports/projects /exports/projects nfs defaults 0 0
```

The status bar at the bottom indicates "3,15 All". The desktop environment includes a dock with various icons and a system tray showing network, battery, and date/time information.

(e) Mount the file and see status, you will see this.



The screenshot shows a terminal window titled "Terminal" with the command "hduser@datanode1: ~". The content of the terminal is as follows:

```
File Edit View Search Terminal Help
Thu 11:30 ●
hduser@datanode1: ~

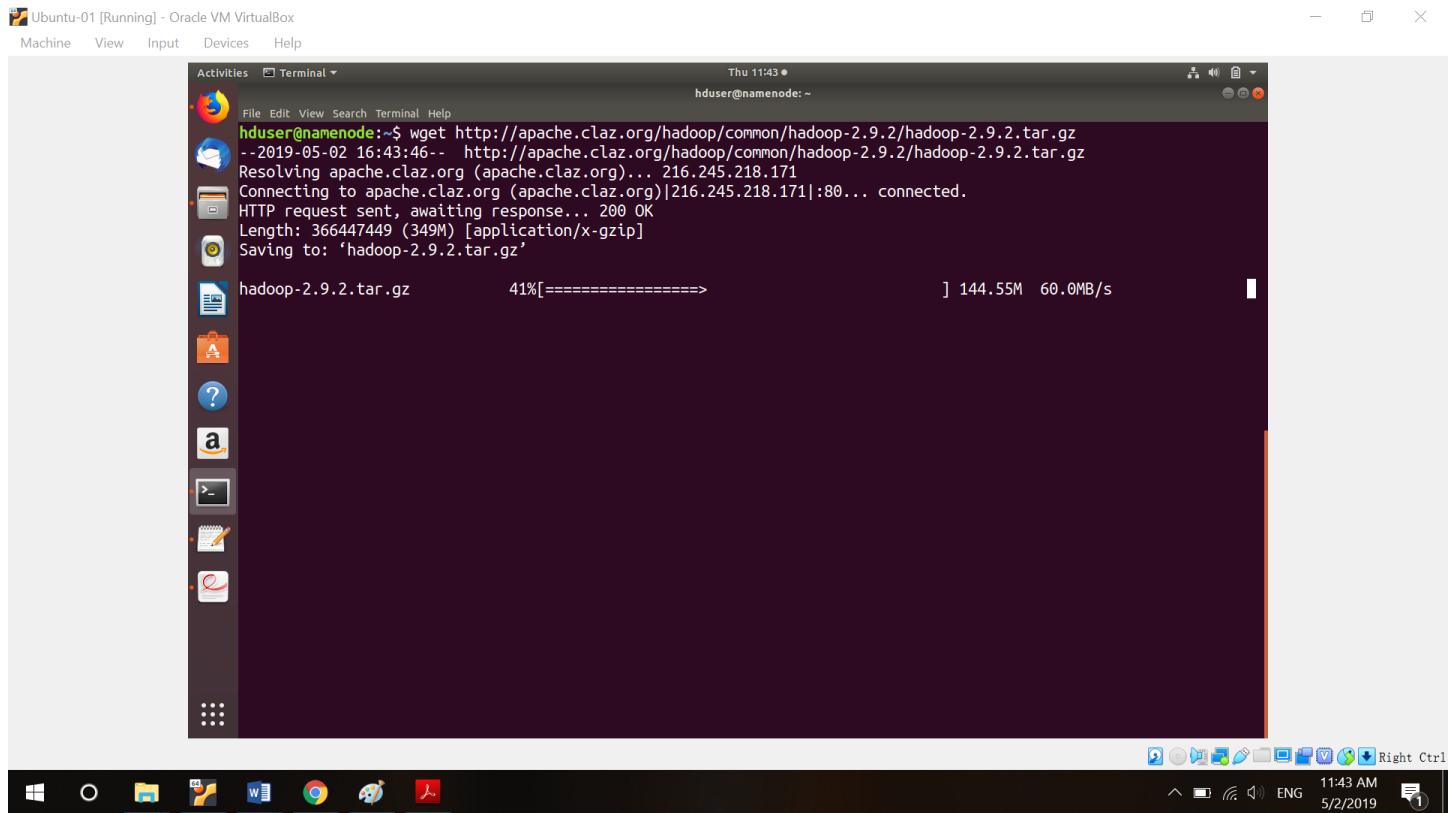
Creating config file /etc/idmapd.conf with new version
Creating config file /etc/default/nfs-common with new version
Adding system user `statd' (UID 112) ...
Adding new user `statd' (UID 112) with group `nogroup' ...
Not creating home directory `/var/lib/nfs'.
nfs-utils.service is a disabled or a static unit, not starting it.
Processing triggers for libc-bin (2.23-0ubuntu11) ...
Processing triggers for ureadahead (0.100.0-19.1) ...
Processing triggers for systemd (229-4ubuntu21.21) ...
hduser@datanode1:~$ sudo vim /etc/idmapd.conf
hduser@datanode1:~$ sudo mkdir -p /exports/projects
hduser@datanode1:~$ sudo chown hduser:hduser /exports/projects
hduser@datanode1:~$ sudo vim /etc/fstab
hduser@datanode1:~$ sudo mount -t nfs \
> 10.154.212.234:/exports/projects /exports/projects
hduser@datanode1:~$ sudo df -hT
Filesystem          Type  Size  Used Avail Use% Mounted on
lxr/containers/datanode1   zfs   48G  599M  47G  2% /
none                tmpfs  492K   0  492K  0% /dev
udev                devtmpfs 63G   0   63G  0% /dev/fuse
tmpfs               tmpfs  100K   0  100K  0% /dev/lxd
tmpfs               tmpfs  100K   0  100K  0% /dev/.lxd-mounts
tmpfs               tmpfs  63G   0   63G  0% /dev/shm
tmpfs               tmpfs  63G  8.3M  63G  1% /run
tmpfs               tmpfs   5.0M   0   5.0M  0% /run/lock
tmpfs               tmpfs  63G   0   63G  0% /sys/fs/cgroup
tmpfs               tmpfs  13G   0   13G  0% /run/user/1001
10.154.212.234:/exports/projects nfs4  48G  599M  47G  2% /exports/projects
hduser@datanode1:~$
```

The status bar at the bottom indicates "11:30 AM 5/2/2019". The desktop environment includes a dock with various icons and a system tray showing network, battery, and date/time information.

(f) Do such thing on all data nodes.

5. Hadoop

(a) Download Hadoop to your system (name node).



Ubuntu-01 [Running] - Oracle VM VirtualBox

Machine View Input Devices Help

Activities Terminal Thu 11:43 •

hduser@namenode:~\$ wget http://apache.claz.org/hadoop/common/hadoop-2.9.2/hadoop-2.9.2.tar.gz

- 2019-05-02 16:43:46-- http://apache.claz.org/hadoop/common/hadoop-2.9.2/hadoop-2.9.2.tar.gz

Resolving apache.claz.org (apache.claz.org)... 216.245.218.171

Connecting to apache.claz.org (apache.claz.org)|216.245.218.171|:80... connected.

HTTP request sent, awaiting response... 200 OK

Length: 366447449 (349M) [application/x-gzip]

Saving to: 'hadoop-2.9.2.tar.gz'

hadoop-2.9.2.tar.gz 41%[=====] 144.55M 60.0MB/s

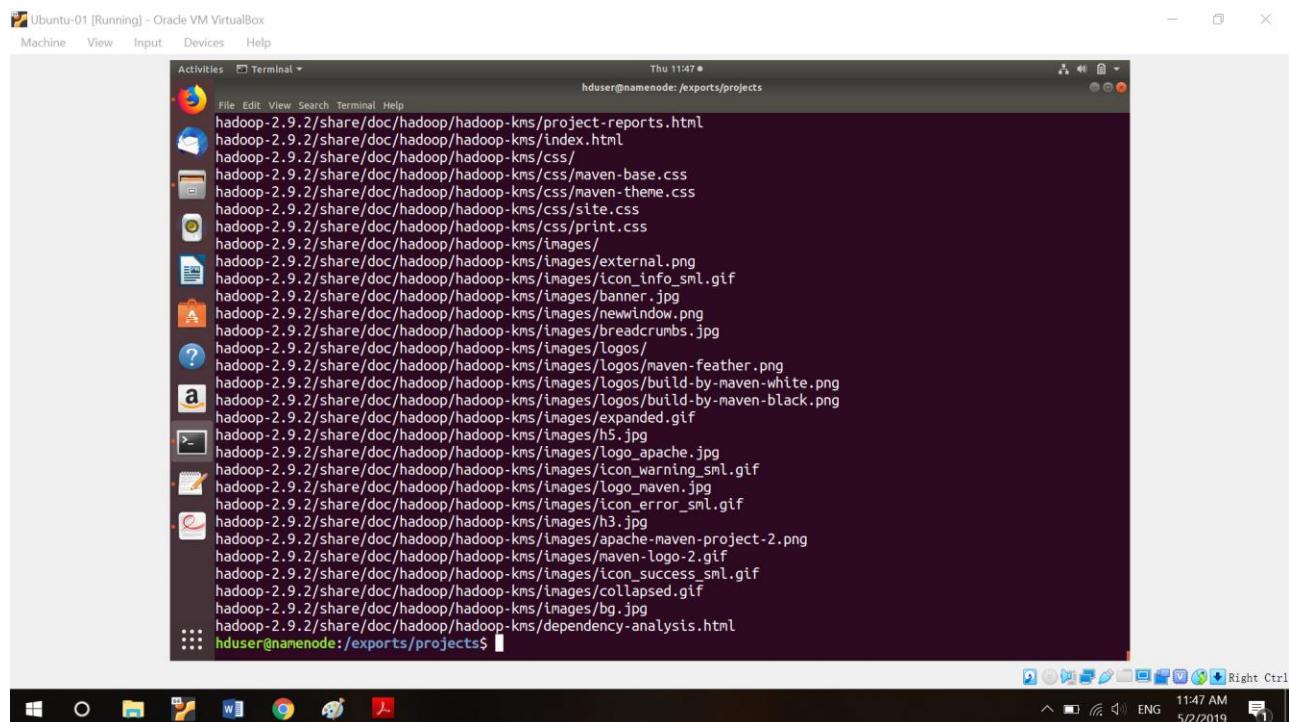
File Edit View Search Terminal Help

hduser@namenode:~\$

Windows Taskbar icons: File Explorer, File History, OneDrive, Mail, Photos, OneNote, Edge, Google Chrome, File Manager, and a red application icon.

System tray: ENG 11:43 AM 5/2/2019

(b) Run such commands “cp hadoop-2.9.2.tar.gz /exports/projects/.”, “cd /exports/projects”, “cd /exports/projects”, “tar xvf hadoop-2.9.2.tar.gz”. Hadoop will be uncompressed.



Ubuntu-01 [Running] - Oracle VM VirtualBox

Machine View Input Devices Help

Activities Terminal Thu 11:47 •

hduser@namenode:/exports/projects\$

hadoop-2.9.2/share/doc/hadoop/hadoop-kms/project-reports.html
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/index.html
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/css/
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/css/maven-base.css
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/css/maven-theme.css
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/css/site.css
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/css/print.css
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/external.png
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/icon_info_sml.gif
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/banner.jpg
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/newindow.png
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/breadcrumbs.jpg
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/logos/
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/Logos/maven-feather.png
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/Logos/build-by-maven-white.png
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/Logos/build-by-maven-black.png
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/expanded.gif
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/h5.jpg
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/logo_apache.jpg
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/icon_warning_sml.gif
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/logo_maven.jpg
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/icon_error_sml.gif
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/h3.jpg
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/apache-maven-project-2.png
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/maven-logo-2.gif
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/icon_success_sml.gif
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/collapsed.gif
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/images/bg.jpg
hadoop-2.9.2/share/doc/hadoop/hadoop-kms/dependency-analysis.html

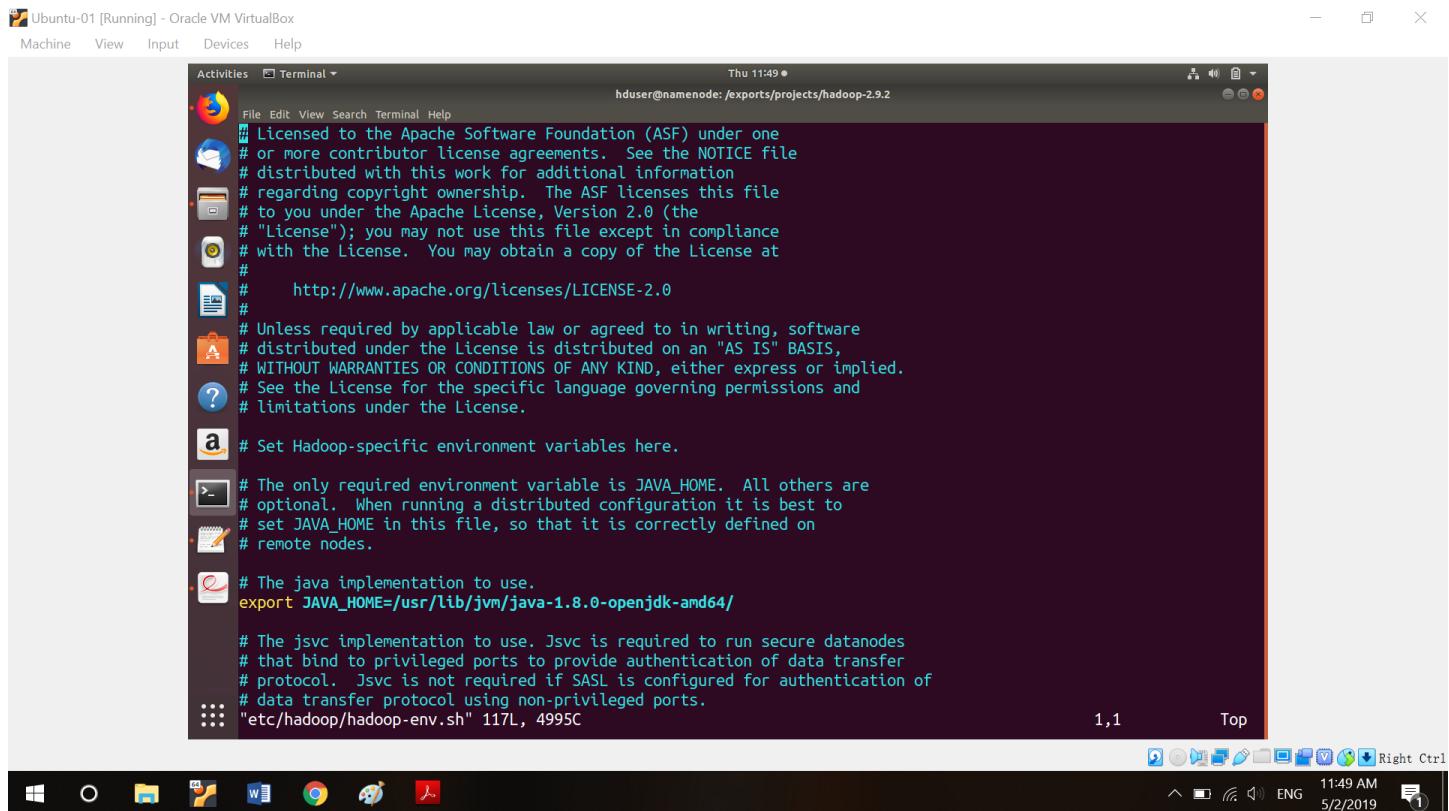
File Edit View Search Terminal Help

hduser@namenode:/exports/projects\$

Windows Taskbar icons: File Explorer, File History, OneDrive, Mail, Photos, OneNote, Edge, Google Chrome, File Manager, and a red application icon.

System tray: ENG 11:47 AM 5/2/2019

(c) Hard code your Java path on Hadoop-env.sh.



The screenshot shows a terminal window titled "Terminal" running on an Ubuntu 01 VM. The window title bar includes "Ubuntu-01 [Running] - Oracle VM VirtualBox" and "Machine View Input Devices Help". The terminal window displays the file "/etc/hadoop/hadoop-env.sh" with the following content:

```
Thu 11:49 ◊
hduser@namenode: /exports/projects/hadoop-2.9.2

# Licensed to the Apache Software Foundation (ASF) under one
# or more contributor license agreements. See the NOTICE file
# distributed with this work for additional information
# regarding copyright ownership. The ASF licenses this file
# to you 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.

# Set Hadoop-specific environment variables here.

# The only required environment variable is JAVA_HOME. All others are
# optional. When running a distributed configuration it is best to
# set JAVA_HOME in this file, so that it is correctly defined on
# remote nodes.

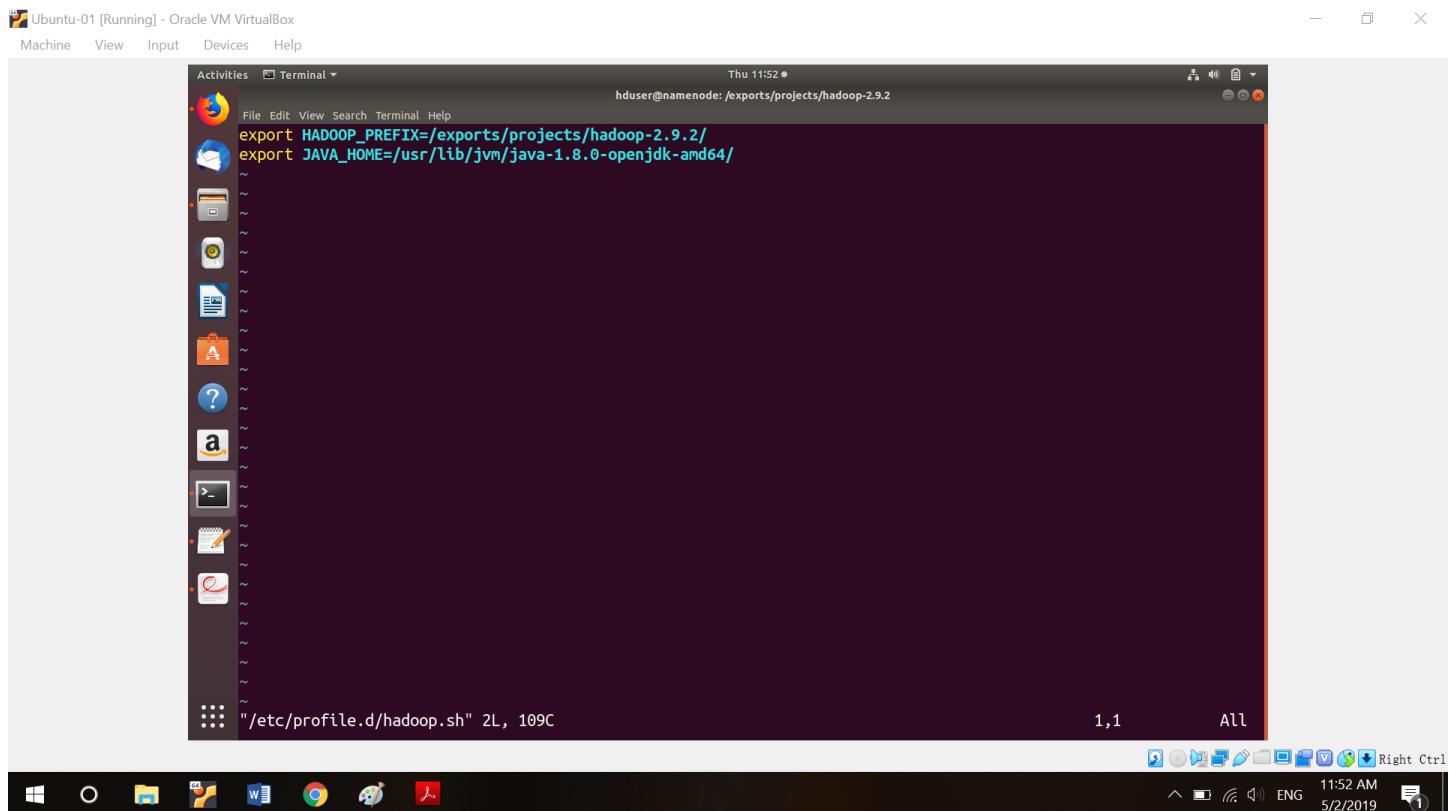
# The java implementation to use.
export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64/

# The jsvc implementation to use. Jsvc is required to run secure datanodes
# that bind to privileged ports to provide authentication of data transfer
# protocol. Jsvc is not required if SASL is configured for authentication of
# data transfer protocol using non-privileged ports.

/etc/hadoop/hadoop-env.sh" 117L, 4995C
```

The status bar at the bottom shows "1,1 Top" and a system tray with icons for battery, signal, and date/time (11:49 AM, 5/2/2019).

(d) On profile.d/Hadoop.sh file, so such thing.



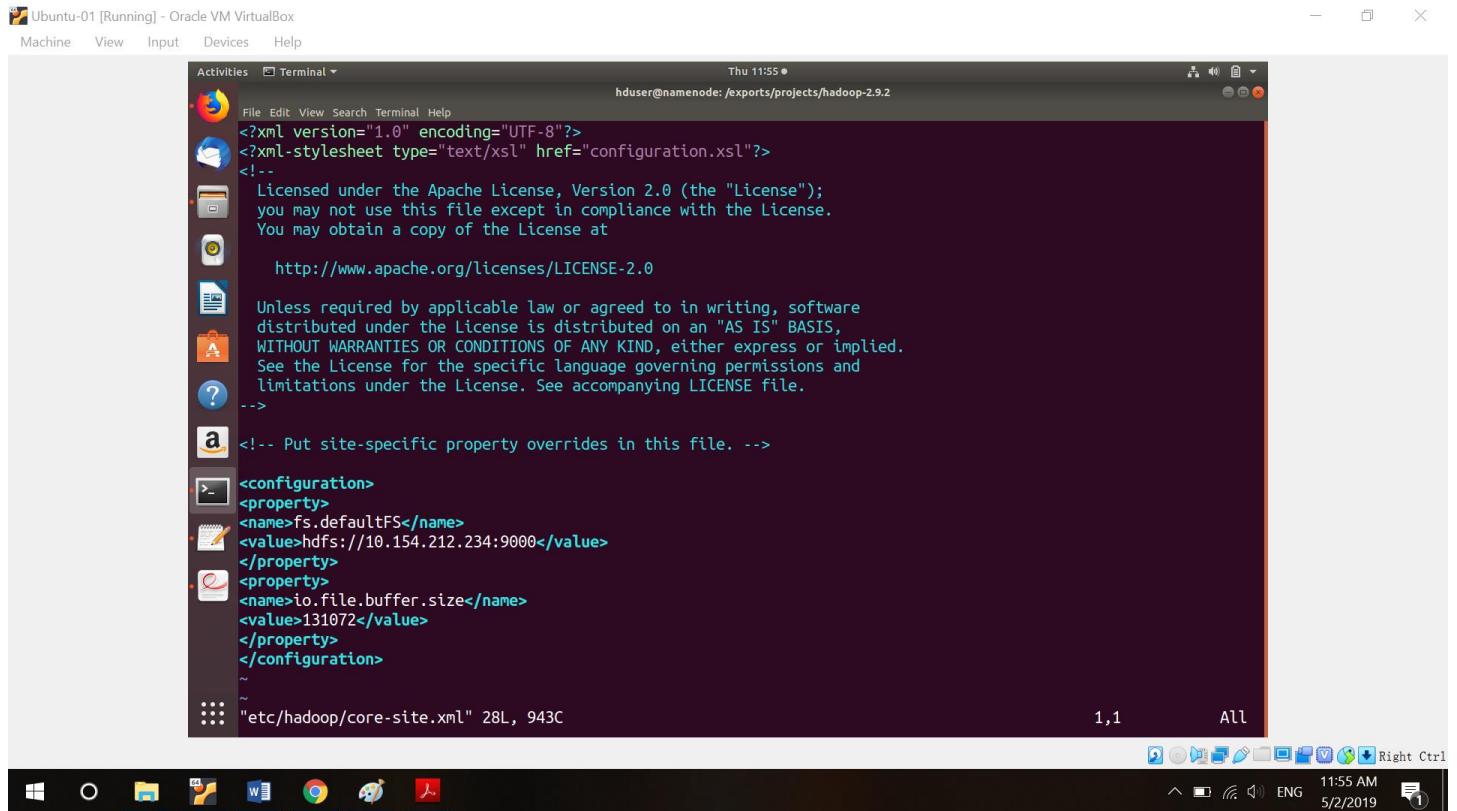
The screenshot shows a terminal window titled "Terminal" running on an Ubuntu 01 VM. The window title bar includes "Ubuntu-01 [Running] - Oracle VM VirtualBox" and "Machine View Input Devices Help". The terminal window displays the file "/etc/profile.d/hadoop.sh" with the following content:

```
Thu 11:52 ◊
hduser@namenode: /exports/projects/hadoop-2.9.2

export HADOOP_PREFIX=/exports/projects/hadoop-2.9.2/
export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64/
~
```

The status bar at the bottom shows "1,1 All" and a system tray with icons for battery, signal, and date/time (11:52 AM, 5/2/2019).

(e) Modify core-site.xml



The screenshot shows a terminal window titled "Terminal" running on an Ubuntu 18.04 LTS desktop. The window title bar includes "Ubuntu-01 [Running] - Oracle VM VirtualBox" and "Machine View Input Devices Help". The terminal window has a dark background with light-colored text. It displays the XML configuration file "core-site.xml" with several properties defined:

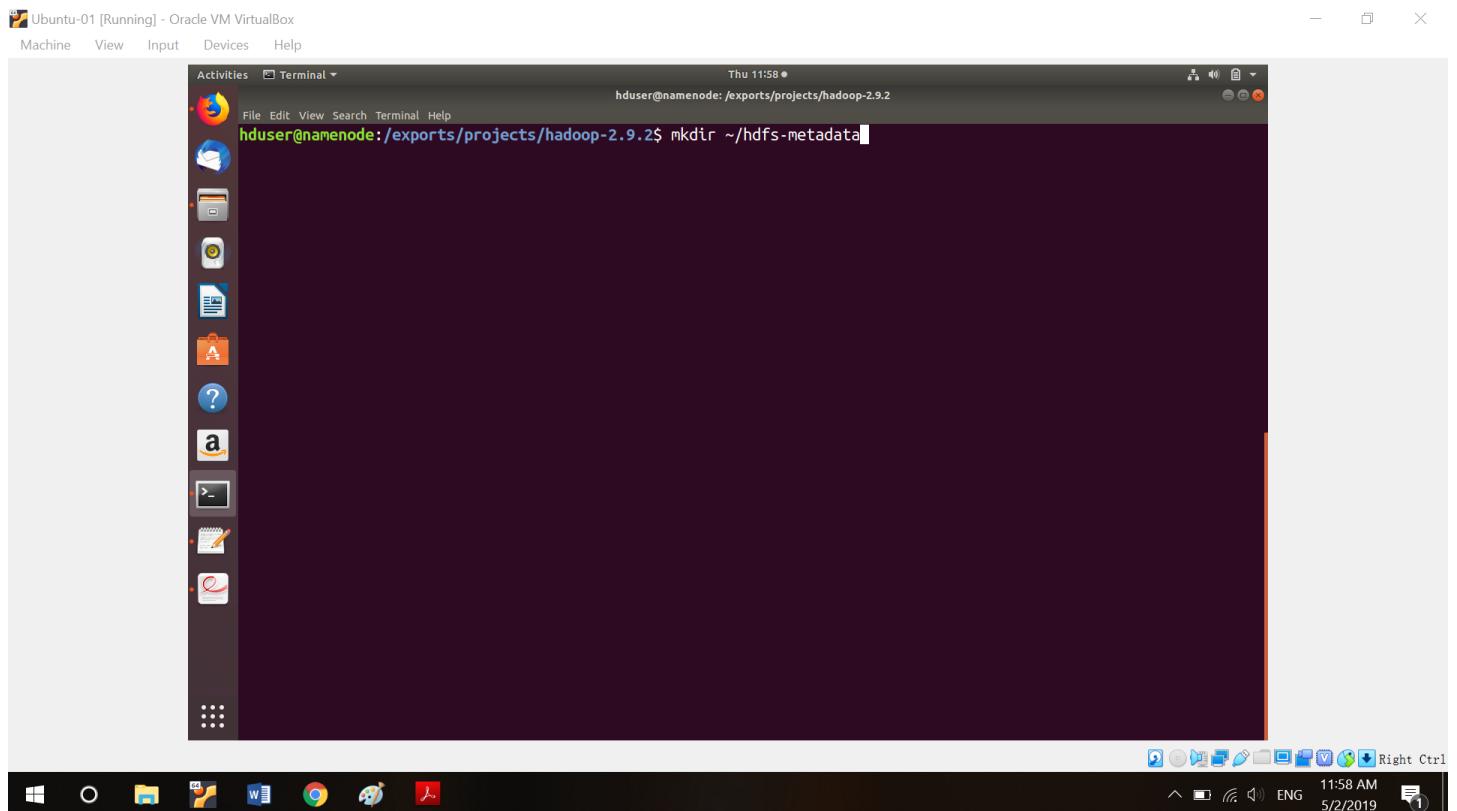
```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

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.defaultFS</name>
<value>hdfs://10.154.212.234:9000</value>
</property>
<property>
<name>io.file.buffer.size</name>
<value>131072</value>
</property>
</configuration>
~>
"etc/hadoop/core-site.xml" 28L, 943C
```

The status bar at the bottom of the terminal window shows "1,1" and "All". The desktop taskbar at the bottom includes icons for File Explorer, Task View, Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Microsoft Edge, and FileZilla. The system tray shows the date and time as "Thu 11:55 AM 5/2/2019" and icons for battery, signal strength, and network.

(d) Create a directory on name node to save your metadata. And create a data directory on your data node.



The screenshot shows a terminal window titled "Terminal" running on an Ubuntu 18.04 LTS desktop. The window title bar includes "Ubuntu-01 [Running] - Oracle VM VirtualBox" and "Machine View Input Devices Help". The terminal window has a dark background with light-colored text. It displays the command "mkdir ~/hdfs-metadata" being typed into the terminal:

```
hduser@namenode:/exports/projects/hadoop-2.9.2$ mkdir ~hdfs-metadata
```

The status bar at the bottom of the terminal window shows "11:58 AM" and "Right Ctrl". The desktop taskbar at the bottom includes icons for File Explorer, Task View, Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Microsoft Edge, and FileZilla. The system tray shows the date and time as "11:58 AM 5/2/2019" and icons for battery, signal strength, and network.

Ubuntu-01 [Running] - Oracle VM VirtualBox

Machine View Input Devices Help

Activities Terminal Thu 12:00 ● hduser@datanode1: ~

```
hduser@namenode:/exports/projects/hadoop-2.9.2$ ssh datanode1
Welcome to Ubuntu 16.04.6 LTS (GNU/Linux 4.4.0-145-generic x86_64)

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

 Get cloud support with Ubuntu Advantage Cloud Guest:
 http://www.ubuntu.com/business/services/cloud

0 packages can be updated.
0 updates are security updates.

? New release '18.04.2 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

a Last login: Thu May 2 16:20:41 2019 from 10.154.212.234
hduser@datanode1:~$ mkdir ~/hduser/hdfs-data
mkdir: cannot create directory '/home/hduser/hduser/hdfs-data': No such file or directory
hduser@datanode1:~$ mkdir ~/hdfs-data
hduser@datanode1:~$
```

Right Ctrl

Windows Start File Explorer Microsoft Word Microsoft Edge Paint Task View

12:00 PM 5/2/2019

(e) Modify hdfs-site.xml file.

Ubuntu-01 [Running] - Oracle VM VirtualBox

Machine View Input Devices Help

Activities Terminal Thu 12:06 ● hduser@namenode:/exports/projects/hadoop-2.9.2

```
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>1</value>
</property>
<property>
<name>dfs.namenode.name.dir</name>
<value>/home/hduser/hdfs-metadata</value>
</property>
<property>
<name>dfs.blocksize</name>
<value>67108864</value>
</property>
<property>
<name>dfs.namenode.handler.count</name>
<value>100</value>
</property>
<property>
<name>dfs.datanode.data.dir</name>
<value>/home/hduser/hdfs-data</value>
</property>
</configuration>
-- INSERT --
```

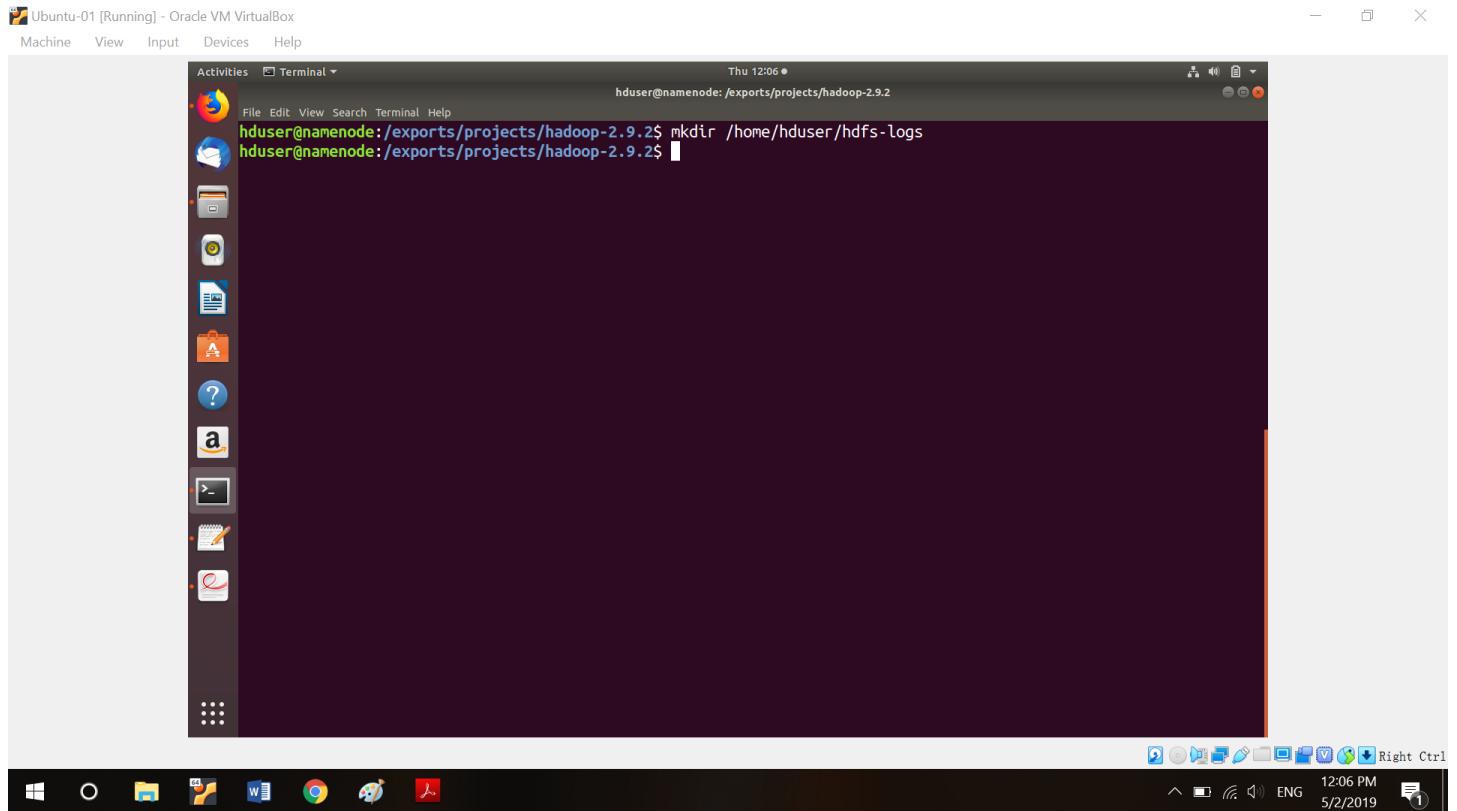
40,17 Bot

Right Ctrl

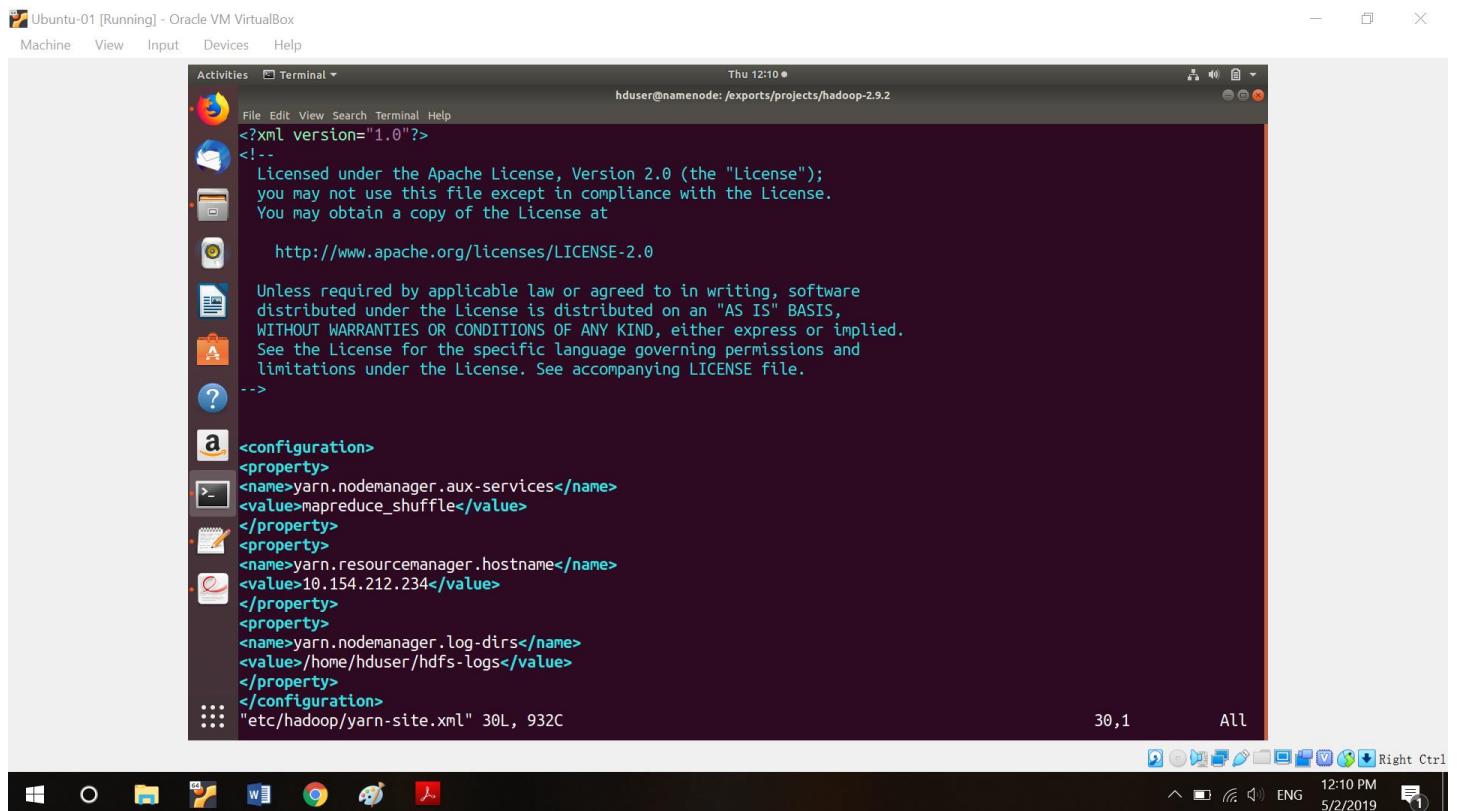
Windows Start File Explorer Microsoft Word Microsoft Edge Paint Task View

12:06 PM 5/2/2019

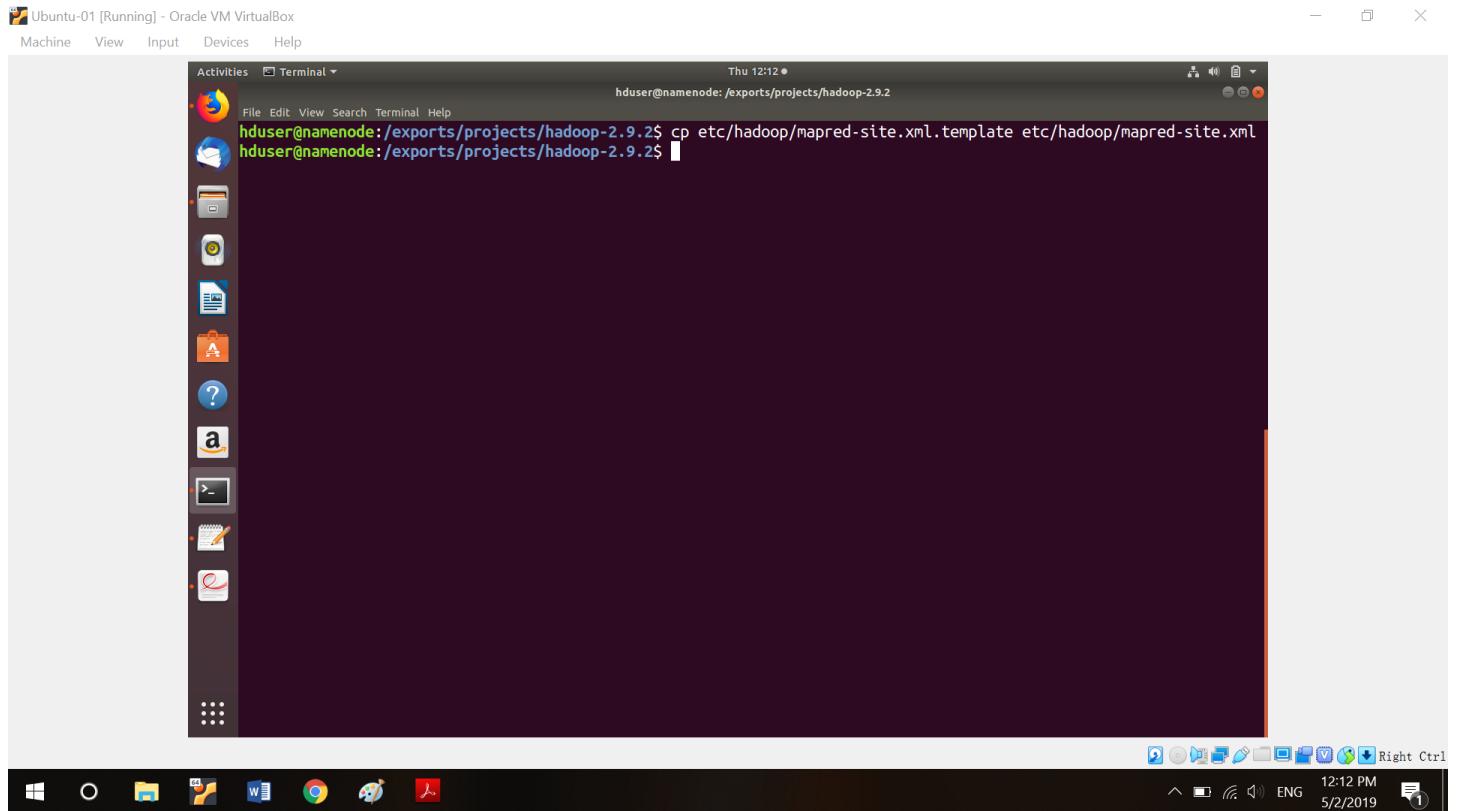
(f) Create log file on all nodes.



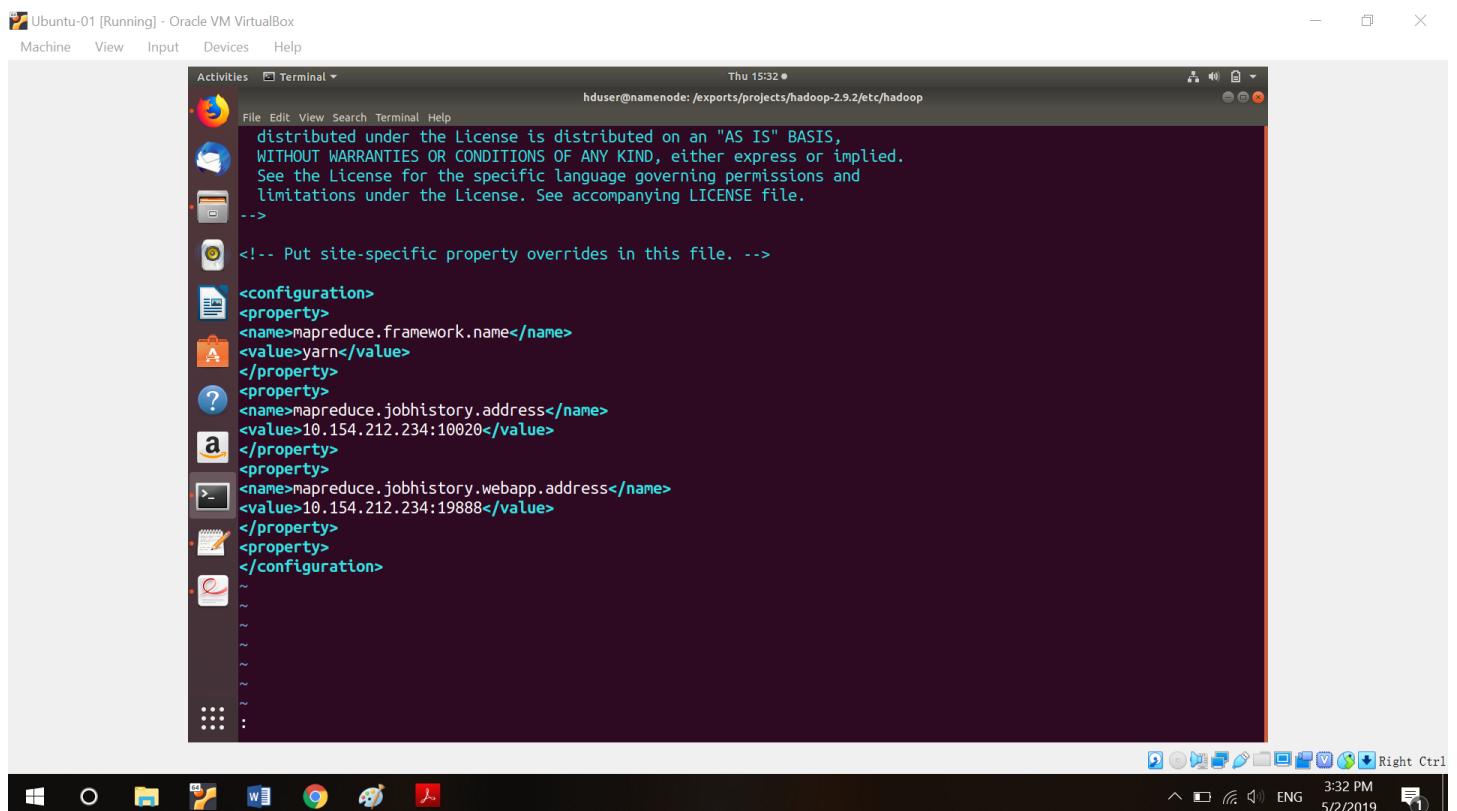
(g) Modify yarn-site.xml.



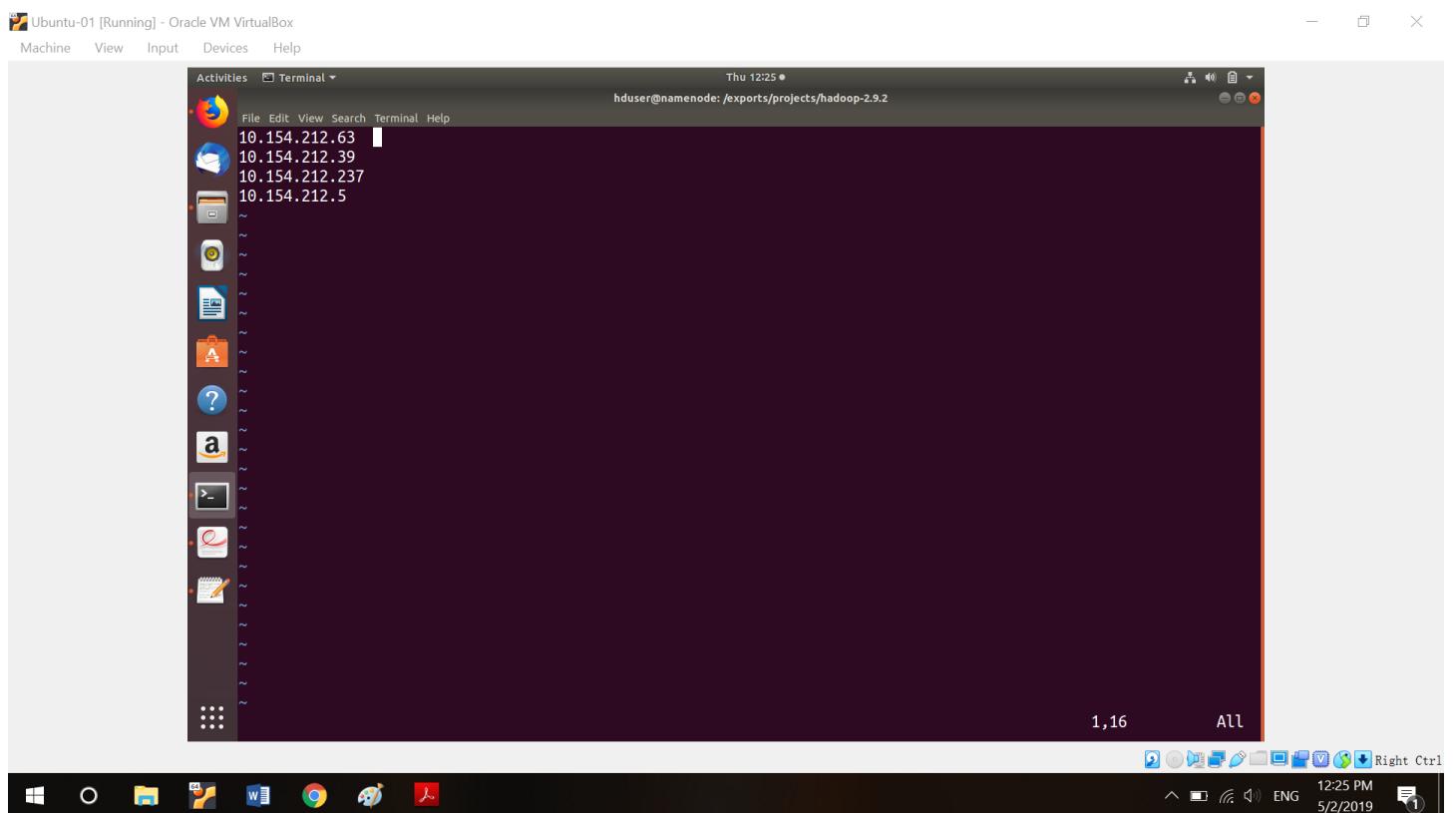
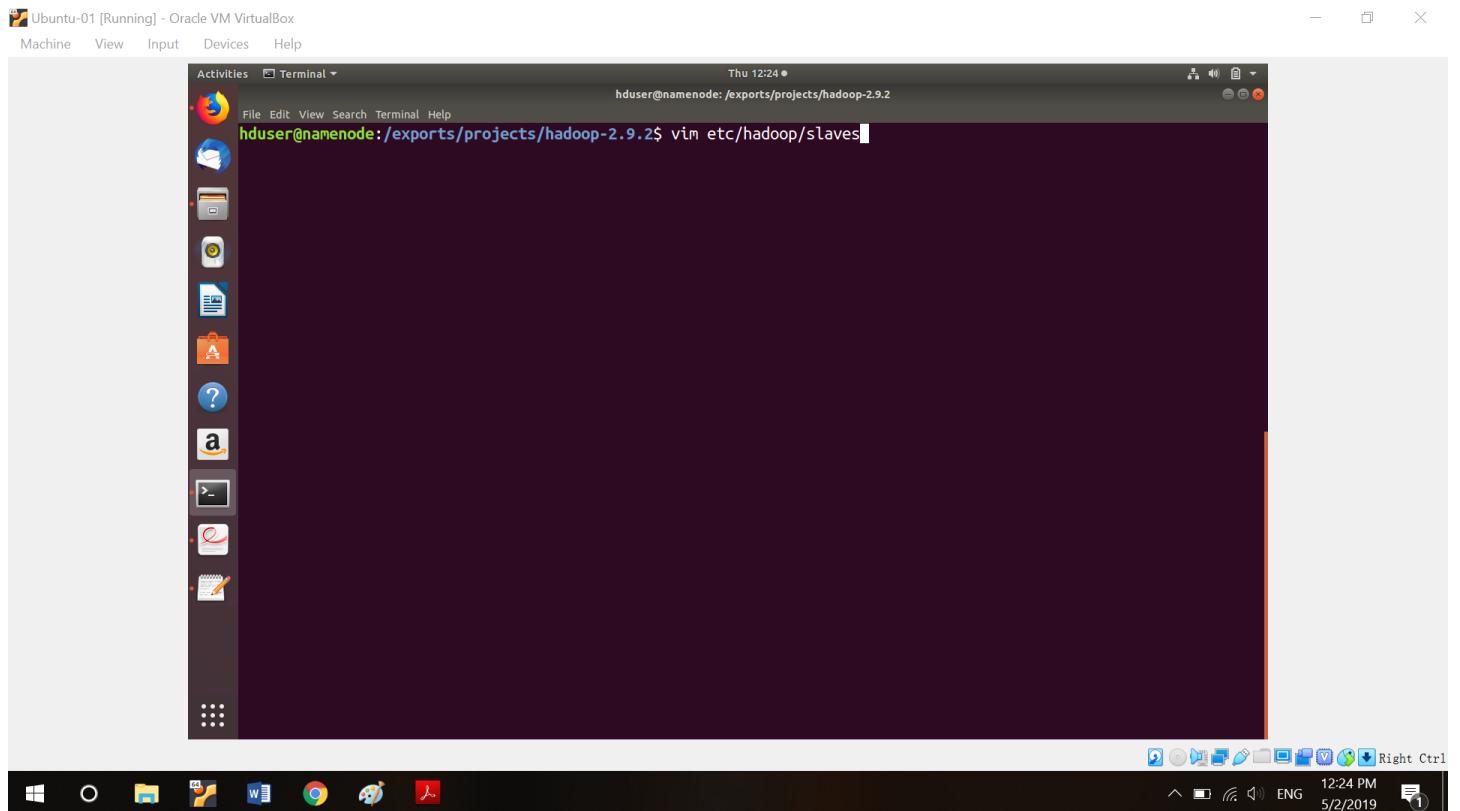
(h) Hadoop higher versions don't contain mapred-site.xml, we should copy it from mapred-site.xml.template.



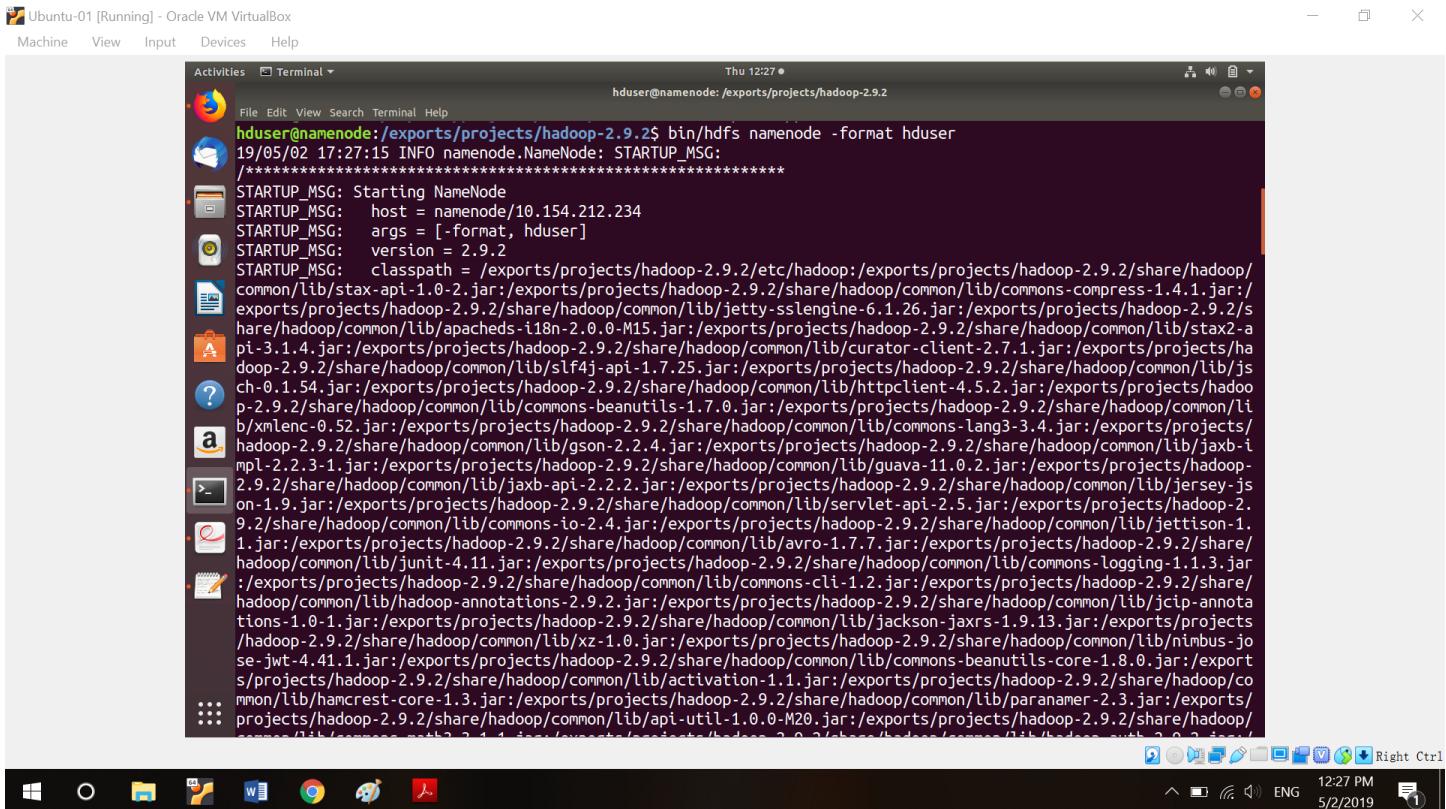
(i) Modify mapred-site.xml. I want to compress the intermediate data because I need to deal with big data.



(j) Tell name node who are data nodes.

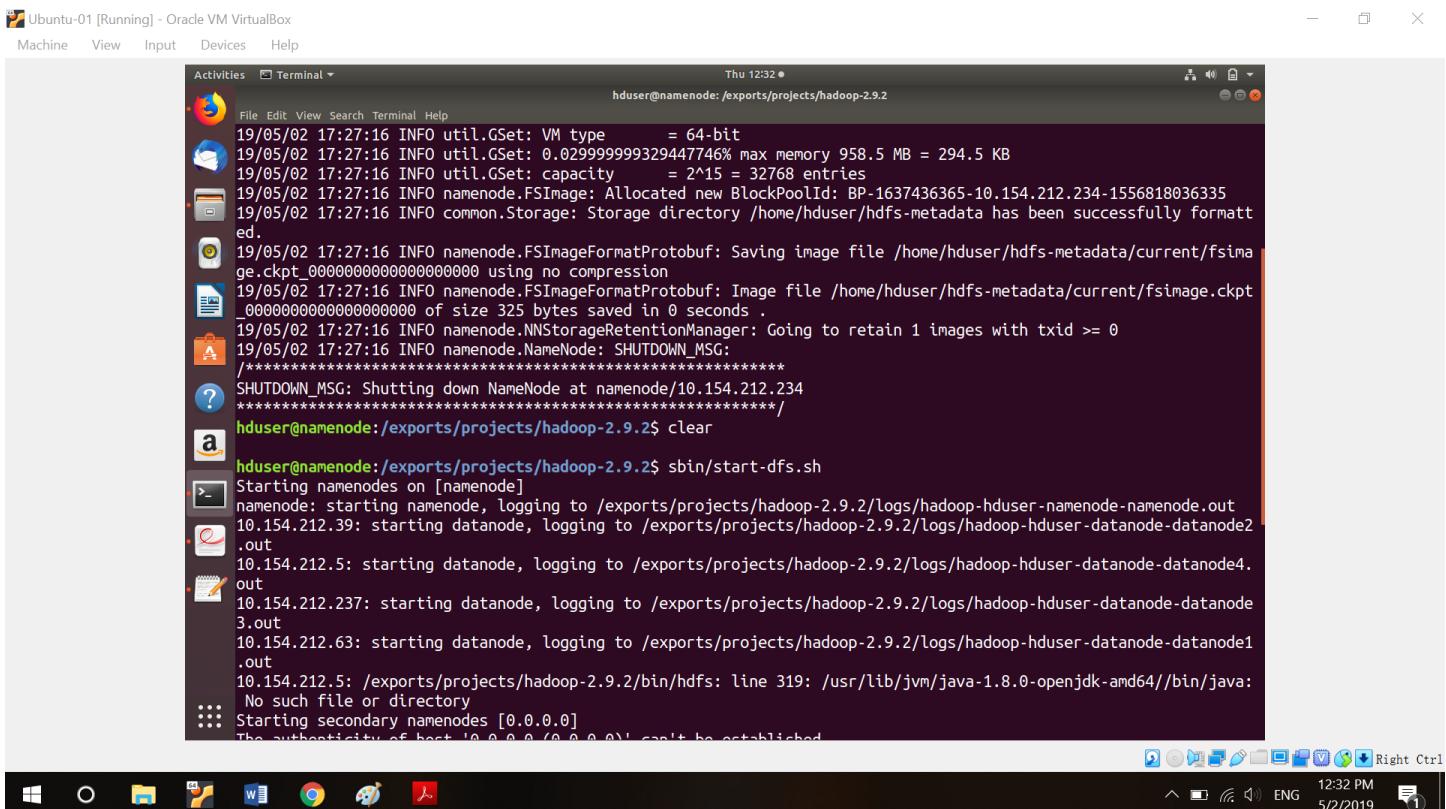


(k) Format and start HDFS, name is hduser.



```
Ubuntu-01 [Running] - Oracle VM VirtualBox
Machine View Input Devices Help
Activities Terminal Thu 12:27
hduser@namenode:/exports/projects/hadoop-2.9.2$ bin/hdfs namenode -format hduser
19/05/02 17:27:15 INFO namenode.NameNode: STARTUP_MSG:
*****STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = namenode/10.154.212.234
STARTUP_MSG: args = [-format, hduser]
STARTUP_MSG: version = 2.9.2
STARTUP_MSG: classpath = /exports/projects/hadoop-2.9.2/etc/hadoop:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/stax-api-1.0-2.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/commons-compress-1.4.1.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/jetty-sslengine-6.1.26.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/apacheds-i18n-2.0.0-M15.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/stax2-a-pi-3.1.4.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/curator-client-2.7.1.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/sl4j-api-1.7.25.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/httpclient-4.5.2.jar:/exports/projects/hadoop-0.1.54.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/httpclient-4.5.2.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/commons-beanutils-1.7.0.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/xmlenc-0.52.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/commons-lang3-3.4.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/gson-2.2.4.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/jaxb-api-2.2.3-1.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/jaxb-api-2.2.2.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/jersey-json-1.9.2.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/servlet-api-2.5.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/jettison-1.1.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/avro-1.7.7.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/junit-4.11.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/commons-logging-1.1.3.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/commons-cli-1.2.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/hadoop-annotations-2.9.2.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/jcip-annotations-1.0-1.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/jackson-jaxrs-1.9.13.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/xz-1.0.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/nimbus-jose-jwt-4.41.1.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/commons-beanutils-core-1.8.0.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/activation-1.1.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/paranamer-2.3.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/api-util-1.0.0-M20.jar:/exports/projects/hadoop-2.9.2/share/hadoop/common/lib/commons-path2-2.1.1.jar:/exports/projects/hadoop-2.9.2/share/hadoop-common/lib/hadoop-path2-2.0.2.jar:/exports/projects/hadoop-2.9.2/share/hadoop-common/lib/hadoop-path2-2.0.2.jar
hduser@namenode:/exports/projects/hadoop-2.9.2$ Right Ctrl
Windows Start File Explorer Microsoft Word Microsoft Edge Paint Task View Taskbar 12:27 PM ENG 5/2/2019
```

(l) Start dfs.



```
Ubuntu-01 [Running] - Oracle VM VirtualBox
Machine View Input Devices Help
Activities Terminal Thu 12:32
hduser@namenode:/exports/projects/hadoop-2.9.2$ Thu 12:32
19/05/02 17:27:16 INFO util.GSet: VM type      = 64-bit
19/05/02 17:27:16 INFO util.GSet: 0.029999999329447746% max memory 958.5 MB = 294.5 KB
19/05/02 17:27:16 INFO util.GSet: capacity      = 2^15 = 32768 entries
19/05/02 17:27:16 INFO namenode.FSImage: Allocated new BlockPoolID: BP-1637436365-10.154.212.234-1556818036335
19/05/02 17:27:16 INFO common.Storage: Storage directory /home/hduser/hdfs-metadata has been successfully formatted.
19/05/02 17:27:16 INFO namenode.FSImageFormatProtobuf: Saving image file /home/hduser/hdfs-metadata/current/fsimage.ckpt_0000000000000000 using no compression
19/05/02 17:27:16 INFO namenode.FSImageFormatProtobuf: Image file /home/hduser/hdfs-metadata/current/fsimage.ckpt_0000000000000000 of size 325 bytes saved in 0 seconds .
19/05/02 17:27:16 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
19/05/02 17:27:16 INFO namenode.NameNode: SHUTDOWN_MSG:
*****SHUTDOWN_MSG: Shutting down NameNode at namenode/10.154.212.234
*****/
hduser@namenode:/exports/projects/hadoop-2.9.2$ clear
hduser@namenode:/exports/projects/hadoop-2.9.2$ sbin/start-dfs.sh
Starting namenodes on [namenode]
namenode: starting namenode, logging to /exports/projects/hadoop-2.9.2/logs/hadoop-hduser-namenode-namenode.out
10.154.212.39: starting datanode, logging to /exports/projects/hadoop-2.9.2/logs/hadoop-hduser-datanode-datanode2.out
10.154.212.5: starting datanode, logging to /exports/projects/hadoop-2.9.2/logs/hadoop-hduser-datanode-datanode4.out
10.154.212.237: starting datanode, logging to /exports/projects/hadoop-2.9.2/logs/hadoop-hduser-datanode-datanode3.out
10.154.212.63: starting datanode, logging to /exports/projects/hadoop-2.9.2/logs/hadoop-hduser-datanode-datanode1.out
10.154.212.5: /exports/projects/hadoop-2.9.2/bin/hdfs: line 319: /usr/lib/jvm/java-1.8.0-openjdk-amd64//bin/java: No such file or directory
Starting secondary namenodes [0.0.0.0]
The authenticity of host '0.0.0.0 (0.0.0.0)' can't be established.
The authenticity of host '0.0.0.0 (0.0.0.0)' can't be established.
hduser@namenode:/exports/projects/hadoop-2.9.2$ Right Ctrl
Windows Start File Explorer Microsoft Word Microsoft Edge Paint Task View Taskbar 12:32 PM ENG 5/2/2019
```

(m) Start yarn.

Hadoop Sort

```
hduser@namenode: /exports/projects/hadoop-2.9.2/bin
File Edit View Search Terminal Help
hduser@namenode:/exports/projects/hadoop-2.9.2/bin$ time ./hadoop jar /exports/projects/hadoop-2.9.2/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.9.2.jar teragen 10000000 /user/input/terasort-input[]
```

```
hduser@namenode: /exports/projects/hadoop-2.9.2/bin$ ./bin/hadoop jar $HADOOP_HOME/share/hadoop/examples/terasort.jar org.apache.hadoop.examples.TeraSort -m 2 -n 1000000000000000000 -D mapreduce.map.memory.mb=1024 -D mapreduce.reduce.memory.mb=1024 -D mapreduce.map.java.opts=-Xms1024m -D mapreduce.reduce.java.opts=-Xms1024m
```

HDFS: Number of write operations=4

Job Counters

- Launched map tasks=2
- Other local map tasks=2
- Total time spent by all maps in occupied slots (ms)=23456
- Total time spent by all reduces in occupied slots (ms)=0
- Total time spent by all map tasks (ms)=23456
- Total vcore-milliseconds taken by all map tasks=23456
- Total megabyte-milliseconds taken by all map tasks=24018944

Map-Reduce Framework

- Map input records=100000000
- Map output records=100000000
- Input split bytes=167
- Spilled Records=0
- Failed Shuffles=0
- Merged Map outputs=0
- GC time elapsed (ms)=390
- CPU time spent (ms)=20480
- Physical memory (bytes) snapshot=422948864
- Virtual memory (bytes) snapshot=3974623232
- Total committed heap usage (bytes)=295698432

org.apache.hadoop.examples.terasort.TeraGen\$Counters

- CHECKSUM=21472776955442690

File Input Format Counters

- Bytes Read=0

File Output Format Counters

- Bytes Written=10000000000

real 0m24.771s

user 0m7.712s

sys 0m0.500s

```
hduser@namenode: /exports/projects/hadoop-2.9.2/bin$ 
```

```
hduser@namenode: /exports/projects/hadoop-2.9.2/bin
File Edit View Search Terminal Help
hduser@namenode:/exports/projects/hadoop-2.9.2/bin$ time ./hadoop jar /exports/projects/hadoop-2.9.2/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.9.2.jar terasort /user/input/terasort-input /user/output/terasort-output
19/05/01 05:08:18 INFO terasort.TeraSort: starting
19/05/01 05:08:19 INFO input.FileInputFormat: Total input files to process : 2
Spent 124ms computing base-splits.
Spent 2ms computing TeraScheduler splits.
Computing input splits took 127ms
Sampling 10 splits of 16
Making 1 from 100000 sampled records
Computing partitions took 1153ms
Spent 1282ms computing partitions.
19/05/01 05:08:20 INFO client.RMProxy: Connecting to ResourceManager at /10.154.212.196:8032
```

```
hduser@namenode: /exports/projects/hadoop-2.9.2/bin
File Edit View Search Terminal Help
Combine input records=0
Combine output records=0
Reduce input groups=10000000
Reduce shuffle bytes=1040000096
Reduce input records=10000000
Reduce output records=10000000
Spilled Records=23355441
Shuffled Maps =16
Failed Shuffles=0
Merged Map outputs=16
GC time elapsed (ms)=2795
CPU time spent (ms)=77100
Physical memory (bytes) snapshot=5088174080
Virtual memory (bytes) snapshot=33554001920
Total committed heap usage (bytes)=3337617408
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=1000000000
File Output Format Counters
Bytes Written=1000000000
19/05/01 05:09:24 INFO terasort.TeraSort: done

real    1m7.398s
user    0m8.696s
sys     0m0.596s
hduser@namenode:/exports/projects/hadoop-2.9.2/bin$
```

```
hduser@namenode: /exports/projects/hadoop-2.9.2/bin
File Edit View Search Terminal Help
hduser@namenode:/exports/projects/hadoop-2.9.2/bin$ time ./hadoop jar /exports/projects/hadoop-2.9.2/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.9.2.jar teravalidate /user/output/terasort -output /user/output/terasort-validate
19/05/01 05:11:13 INFO client.RMProxy: Connecting to ResourceManager at /10.154.212.196:8032
19/05/01 05:11:14 INFO input.FileInputFormat: Total input files to process : 1
Spent 44ms computing base-splits.
Spent 6ms computing TeraScheduler splits.
19/05/01 05:11:15 INFO mapreduce.JobSubmitter: number of splits:1
19/05/01 05:11:15 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
19/05/01 05:11:15 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1556676145780_0007
19/05/01 05:11:15 INFO impl.YarnClientImpl: Submitted application application_1556676145780_0007
19/05/01 05:11:15 INFO mapreduce.Job: The url to track the job: http://namenode:8088/proxy/application_1556676145780_0007/
19/05/01 05:11:15 INFO mapreduce.Job: Running job: job_1556676145780_0007
```

```
hduser@namenode: /exports/projects/hadoop-2.9.2/bin
File Edit View Search Terminal Help
Input split bytes=132
Combine input records=0
Combine output records=0
Reduce input groups=3
Reduce shuffle bytes=94
Reduce input records=3
Reduce output records=1
Spilled Records=6
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=368
CPU time spent (ms)=9840
Physical memory (bytes) snapshot=483086336
Virtual memory (bytes) snapshot=3956531200
Total committed heap usage (bytes)=348127232
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=1000000000
File Output Format Counters
Bytes Written=24

real    0m28.368s
user    0m6.912s
sys     0m0.552s
hduser@namenode:/exports/projects/hadoop-2.9.2/bin$
```

```
hduser@namenode: ~/terasort-validate
File Edit View Search Terminal Help
hduser@namenode:~$ ./hadoop fs -get /user/output/terasort-validate ~/
hduser@namenode:~$ ./hadoop: No such file or directory
hduser@namenode:~$ ls
hduser@namenode:~/hadoop-2.9.2.tar.gz hdfs-data hdfs-logs hdfs-metadata terasort-validate
hduser@namenode:~$ cd terasort-validate/
hduser@namenode:~/terasort-validate$ ls
part-r-00000 _SUCCESS
hduser@namenode:~/terasort-validate$
```

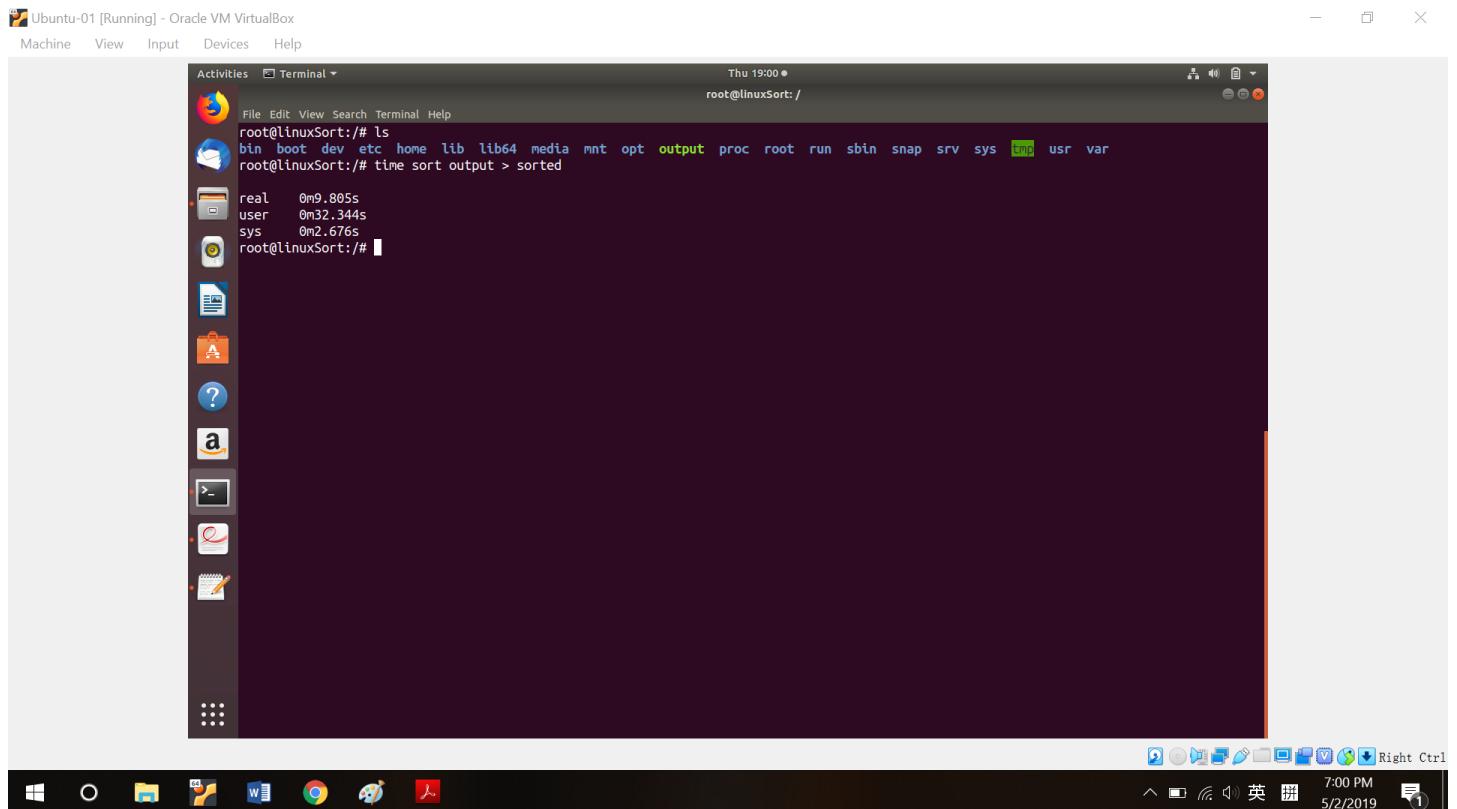
```
hduser@namenode: /exports/projects/hadoop-2.9.2/bin
File Edit View Search Terminal Help
Job Counters
    Launched map tasks=2
    Other local map tasks=2
    Total time spent by all maps in occupied slots (ms)=1021907
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=1021907
    Total vcore-milliseconds taken by all map tasks=1021907
    Total megabyte-milliseconds taken by all map tasks=1046432768
Map-Reduce Framework
    Map input records=6000000000
    Map output records=6000000000
    Input split bytes=170
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=6264
    CPU time spent (ms)=839990
    Physical memory (bytes) snapshot=670027776
    Virtual memory (bytes) snapshot=3957772288
    Total committed heap usage (bytes)=418381824
org.apache.hadoop.examples.terasort.TeraGen$Counters
    CHECKSUM=1288513213283008101
File Input Format Counters
    Bytes Read=0
File Output Format Counters
    Bytes Written=600000000000
real    8m44.533s
user    0m8.936s
sys     0m0.900s
hduser@namenode:/exports/projects/hadoop-2.9.2/bin$
```

```
hduser@namenode: /exports/projects/hadoop-2.9.2/bin
File Edit View Search Terminal Help
19/05/01 05:34:29 INFO mapreduce.Job: map 72% reduce 23%
19/05/01 05:34:30 INFO mapreduce.Job: map 73% reduce 23%
19/05/01 05:34:31 INFO mapreduce.Job: map 73% reduce 24%
19/05/01 05:34:36 INFO mapreduce.Job: map 74% reduce 24%
19/05/01 05:34:37 INFO mapreduce.Job: map 75% reduce 24%
19/05/01 05:34:39 INFO mapreduce.Job: map 76% reduce 24%
19/05/01 05:34:40 INFO mapreduce.Job: map 77% reduce 24%
19/05/01 05:34:41 INFO mapreduce.Job: map 78% reduce 24%
19/05/01 05:34:43 INFO mapreduce.Job: map 78% reduce 25%
19/05/01 05:34:44 INFO mapreduce.Job: map 79% reduce 25%
19/05/01 05:34:50 INFO mapreduce.Job: map 80% reduce 25%
19/05/01 05:34:56 INFO mapreduce.Job: map 81% reduce 25%
19/05/01 05:35:01 INFO mapreduce.Job: map 82% reduce 27%
19/05/01 05:35:02 INFO mapreduce.Job: map 83% reduce 27%
19/05/01 05:35:04 INFO mapreduce.Job: map 84% reduce 27%
19/05/01 05:35:05 INFO mapreduce.Job: map 85% reduce 27%
19/05/01 05:35:07 INFO mapreduce.Job: map 85% reduce 28%
19/05/01 05:35:09 INFO mapreduce.Job: map 86% reduce 28%
19/05/01 05:35:13 INFO mapreduce.Job: map 86% reduce 29%
19/05/01 05:35:15 INFO mapreduce.Job: map 87% reduce 29%
19/05/01 05:35:21 INFO mapreduce.Job: map 88% reduce 29%
19/05/01 05:35:22 INFO mapreduce.Job: map 89% reduce 29%
19/05/01 05:35:24 INFO mapreduce.Job: map 92% reduce 29%
19/05/01 05:35:31 INFO mapreduce.Job: map 93% reduce 29%
19/05/01 05:35:35 INFO mapreduce.Job: map 95% reduce 29%
19/05/01 05:35:36 INFO mapreduce.Job: map 95% reduce 30%
19/05/01 05:35:42 INFO mapreduce.Job: map 96% reduce 32%
19/05/01 05:35:50 INFO mapreduce.Job: map 98% reduce 32%
19/05/01 05:35:51 INFO mapreduce.Job: map 99% reduce 32%
19/05/01 05:35:54 INFO mapreduce.Job: map 99% reduce 33%
19/05/01 05:35:56 INFO mapreduce.Job: map 100% reduce 33%
```

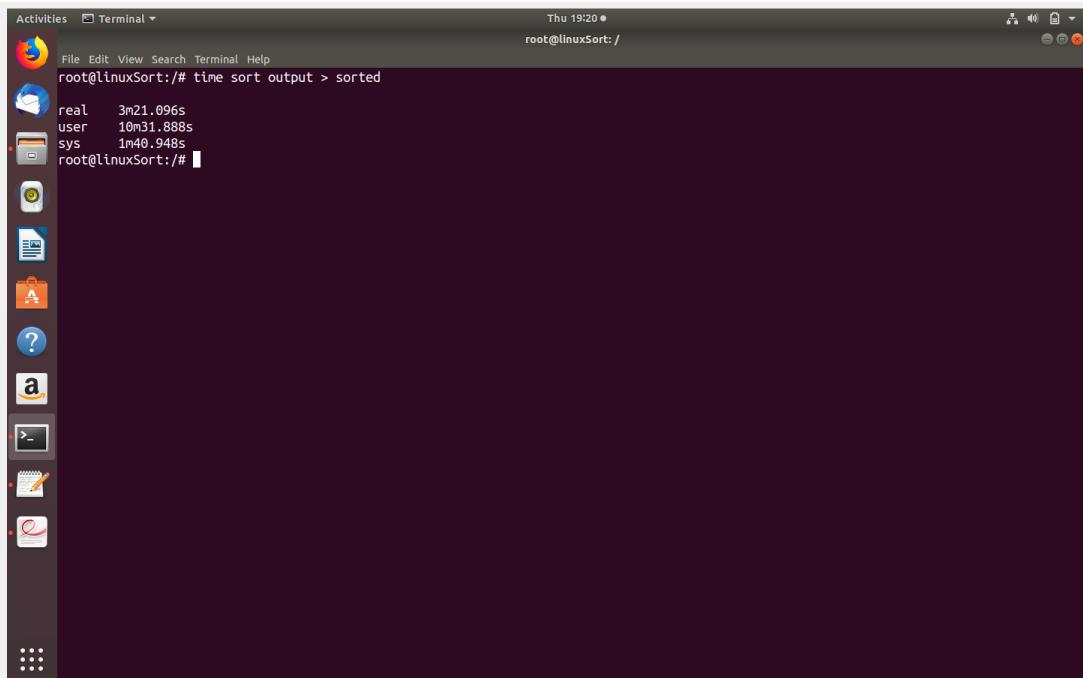
```
hduser@namenode: /exports/projects/hadoop-2.9.2/bin
File Edit View Search Terminal Help
Combine input records=0
Combine output records=0
Reduce input groups=150000000
Reduce shuffle bytes=15600001344
Reduce input records=150000000
Reduce output records=150000000
Spilled Records=542448297
Shuffled Maps =224
Failed Shuffles=0
Merged Map outputs=224
GC time elapsed (ms)=40528
CPU time spent (ms)=1309500
Physical memory (bytes) snapshot=66145480704
Virtual memory (bytes) snapshot=443869577216
Total committed heap usage (bytes)=44228935680
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=15000000000
File Output Format Counters
Bytes Written=15000000000
19/05/01 05:42:00 INFO terasort.TeraSort: done
real    11m28.567s
user    0m11.896s
sys     0m1.024s
hduser@namenode:/exports/projects/hadoop-2.9.2/bin$
```

```
hduser@namenode: /exports/projects/hadoop-2.9.2/bin
File Edit View Search Terminal Help
HDFS: Number of write operations=4
Job Counters
    Launched map tasks=2
    Other local map tasks=2
    Total time spent by all maps in occupied slots (ms)=1093541
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=1093541
    Total vcore-milliseconds taken by all map tasks=1093541
    Total megabyte-milliseconds taken by all map tasks=1119785984
Map-Reduce Framework
    Map input records=6000000000
    Map output records=6000000000
    Input split bytes=170
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=5654
    CPU time spent (ms)=838230
    Physical memory (bytes) snapshot=731353088
    Virtual memory (bytes) snapshot=4032106496
    Total committed heap usage (bytes)=417333248
org.apache.hadoop.examples.terasort.TeraGen$Counters
    CHECKSUM=1288513213283008101
File Input Format Counters
    Bytes Read=0
File Output Format Counters
    Bytes Written=600000000000

real    9m19.099s
user    0m10.132s
sys     0m0.956s
hduser@namenode:/exports/projects/hadoop-2.9.2/bin$
```

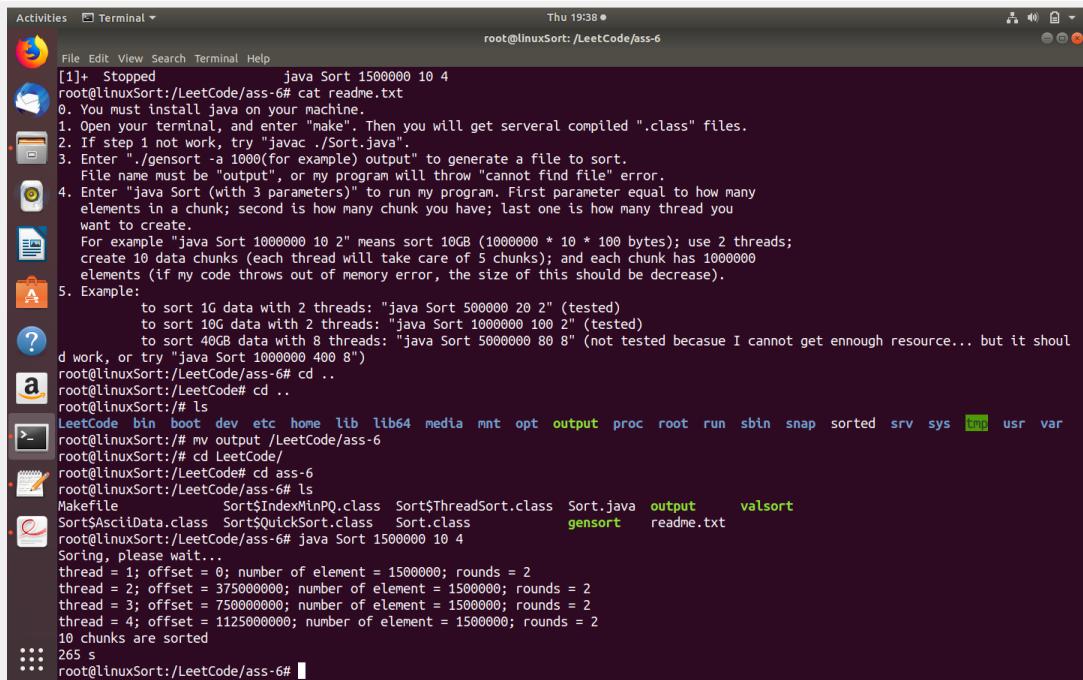


Machine View Input Devices Help



Right Ctrl
7:38 PM
5/2/2019

Machine View Input Devices Help



Right Ctrl
7:38 PM
5/2/2019

```
Activities Terminal Thu 20:09 ●  
cc@llu25-hw8: ~/LeetCode/ass-6  
cc@llu25-hw8:~/LeetCode/ass-6$ java Sort 1000000 10 4  
Sorting, please wait...  
thread = 1; offset = 0; number of element = 1000000; rounds = 2  
thread = 2; offset = 250000000; number of element = 1000000; rounds = 2  
thread = 3; offset = 500000000; number of element = 1000000; rounds = 2  
thread = 4; offset = 750000000; number of element = 1000000; rounds = 2  
10 chunks are sorted  
87 s  
cc@llu25-hw8:~/LeetCode/ass-6$
```



```
Activities Terminal Thu 20:28 ●  
root@sort:/LeetCode/ass-6  
root@sort:/LeetCode/ass-6# time sort output > sorted  
real    0m10.170s  
user    0m32.572s  
sys     0m2.852s  
root@sort:/LeetCode/ass-6#
```



Thu 20:58 ● cc@llu25-hw8: ~/LeetCode/ass-6

```
cc@llu25-hw8:~/LeetCode/ass-6$ ./gensort -a 150000000 output
cc@llu25-hw8:~/LeetCode/ass-6$ java Sort 15000000 10 16
Soring, please wait...
thread = 1; offset = 0; number of element = 15000000; rounds = 0
thread = 2; offset = 937500000; number of element = 15000000; rounds = 0
thread = 3; offset = 1875000000; number of element = 15000000; rounds = 0
thread = 4; offset = 2812500000; number of element = 15000000; rounds = 0
thread = 5; offset = 3750000000; number of element = 15000000; rounds = 0
thread = 6; offset = 4687500000; number of element = 15000000; rounds = 0
thread = 7; offset = 5625000000; number of element = 15000000; rounds = 0
thread = 8; offset = 6562500000; number of element = 15000000; rounds = 0
thread = 9; offset = 7500000000; number of element = 15000000; rounds = 0
thread = 10; offset = 8437500000; number of element = 15000000; rounds = 0
thread = 11; offset = 9375000000; number of element = 15000000; rounds = 0
thread = 12; offset = 10312500000; number of element = 15000000; rounds = 0
thread = 13; offset = 11250000000; number of element = 15000000; rounds = 0
thread = 14; offset = 12187500000; number of element = 15000000; rounds = 0
thread = 15; offset = 13125000000; number of element = 15000000; rounds = 0
thread = 16; offset = 14062500000; number of element = 15000000; rounds = 0
10 chunks are sorted
1371 s
cc@llu25-hw8:~/LeetCode/ass-6$
```

Thu 21:26 ● root@sort:/LeetCode/ass-6

```
root@sort:/LeetCode/ass-6# ls
Makefile           Sort$QuickSort.class   Sort.java  readme.txt
Sort$AsciiData.class  Sort$ThreadSort.class  gensort    valsrt
Sort$IndexMinPQ.class  Sort.class          output
root@sort:/LeetCode/ass-6# time sort output > sort
real    5m44.355s
user    11m1.780s
sys     1m48.752s
root@sort:/LeetCode/ass-6#
```

Experiment	Shred Memory	Linux	Hadoop Sort	Spark Sort
1 small, 1G	87s	9.8s	-	-
1 small, 15G	265s	201s	-	-
1 large, 1G	65s	44s		
1 large, 15G	1371s	661s		
1 large, 60G	5498s	2735s		
4 small, 1G	-	-	67s	89s
4 small, 15G	-	-	525s	934s
4 small, 60G	-	-	2486s	4856s