Install Virtual Machines

We will use 4 VM with differentes packages:

- **elephant**: NameNode, DataNode, TaskTracker.
- **tiger**: SecundaryNamenode, DataNode and TaskTracker.
- **horse**: JobTracker, DataNode and TaskTracker.
- **monkey**: DataNode and TaskTracker.

First, we configure one VM with basic components and network configurations. Later, we copy for 3 others VMs.

1. Configure keyboard
   System -> Keyboard. Go to Layouts -> Add... -> Country -> *Spain*

See: https://access.redhat.com/documentation/en-US/Red_Hat_Enterprise_Linux/3/html/System_Administration_Guide/ch-keyboardconfig.html

   Select Spain as default and remove US keyboard.

2. Set Clock settings

Sudoer user password by default is "*cloudera*".

3. Update packages

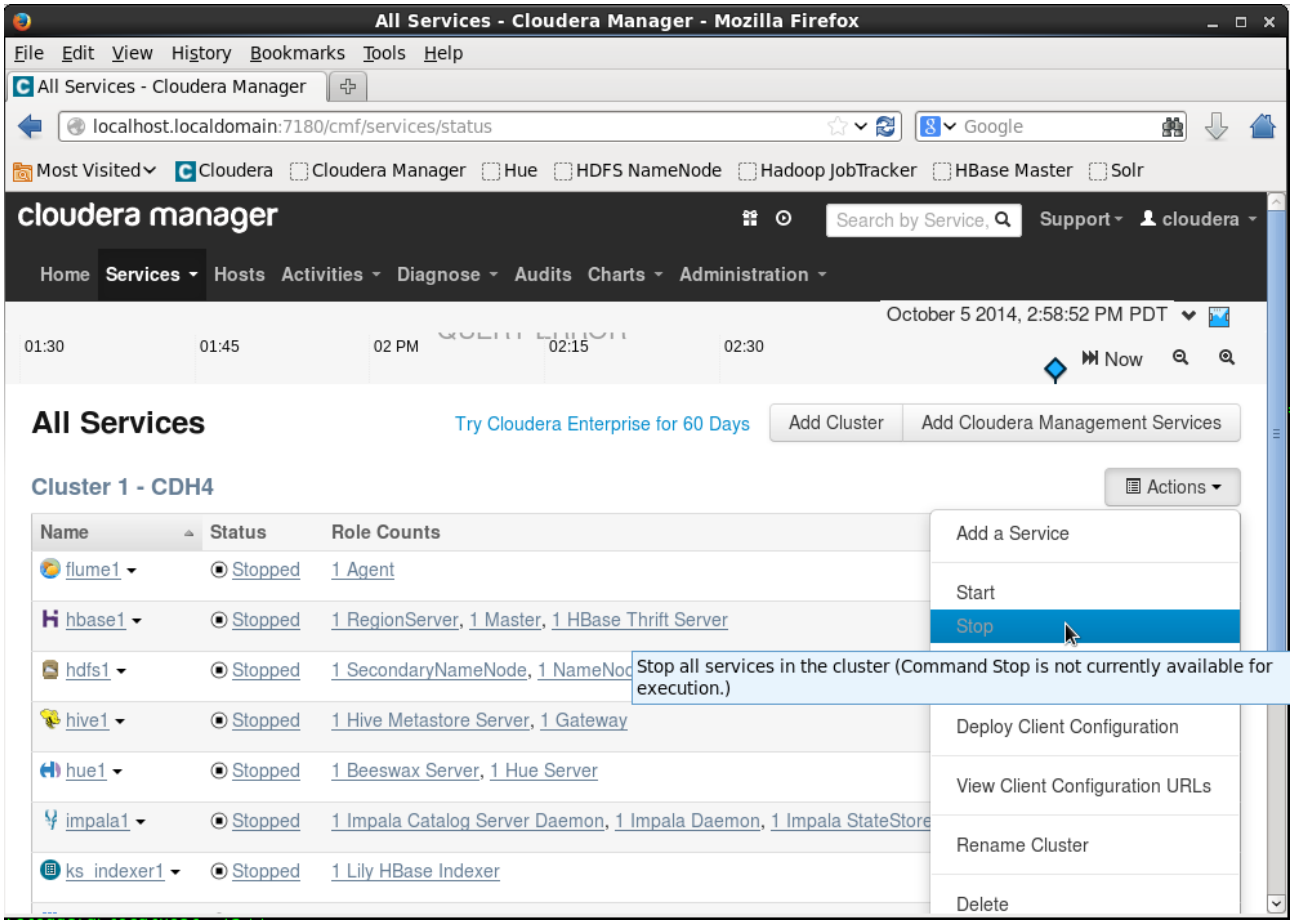
$ yum update -y

If you want a vim editor

$ yum install vim -y

4. Stop Cloudera Manager

- Go to file:///home/cloudera/Documents/quick-hadoop/cloudera-manager.html
- Click on Actions -> Stop



- Stop Clouder Services:
  one-step-one
  - for s in `cd /etc/init.d/; ls cloudera*`; do sudo service $s stop; done

  or serveral-steps
  - $ sudo service cloudera-scm-server stop
  - $ sudo service cloudera-scm-server-db stop
  - $ sudo service cloudera-scm-agent hard_stop
- Remove from start on boot
  - $ sudo chkconfig --list|grep cloudera
  - $ for s in `cd /etc/init.d/; ls cloudera*`; do sudo chkconfig $s on; done

5. Remove pseudodistributed configuration

- $ sudo yum remove -y hadoop-0.20-conf-pseudo

6. Reinstall basic packages

- $ sudo yum -y reinstall hadoop-hdfs-namenode hadoop-hdfs-datanode hadoop-0.20-mapreduce-tasktracker hadoop-hdfs-secondarynamenode hadoop-0.20-mapreduce-jobtracker
- $ for s in `cd /etc/init.d/; ls hadoop*`; do sudo chkconfig $s on; done

7. Configure network

- Configure dhcp with Network-Manager for eth1 (we will use NAT config with guests OS):

  $ sudo vim /etc/sysconfig/network-scripts/ifcfg-eth1

DEVICE="eth1"
TYPE=Ethernet
BOOTPROTO="dhcp"
NM_CONTROLLED="yes"
DEFROUTE=yes
IPV4_FAILURE_FATAL=yes
IPV6INIT=no
NAME="eth1"
ONBOOT=yes

- Configure static IP on eth2 (Internal Network device)
  $ sudo nano /etc/sysconfig/network-scripts/ifcfg-eth2
  DEVICE="eth2"
  TYPE=Ethernet
  BOOTPROTO="static"
  NM_CONTROLLED="no"
  IPADDR=192.168.1.1
  PREFIX=24
  DEFROUTE=yes
  IPV4_FAILURE_FATAL=yes
  IPV6INIT=no
  NAME="eth2"
  ONBOOT=yes

- Configure hostname
  $ sudo nano /etc/sysconfig/network
  NETWORKING=yes
  HOSTNAME=**elephant**

- Configure hosts
  - 4 VMs

  $ sudo nano **/etc/hosts**
  192.168.1.1 elephant
  192.168.1.2 tiger
  192.168.1.3 horse
  192.168.1.4 monkey

- Restart network
  $ sudo service network restart

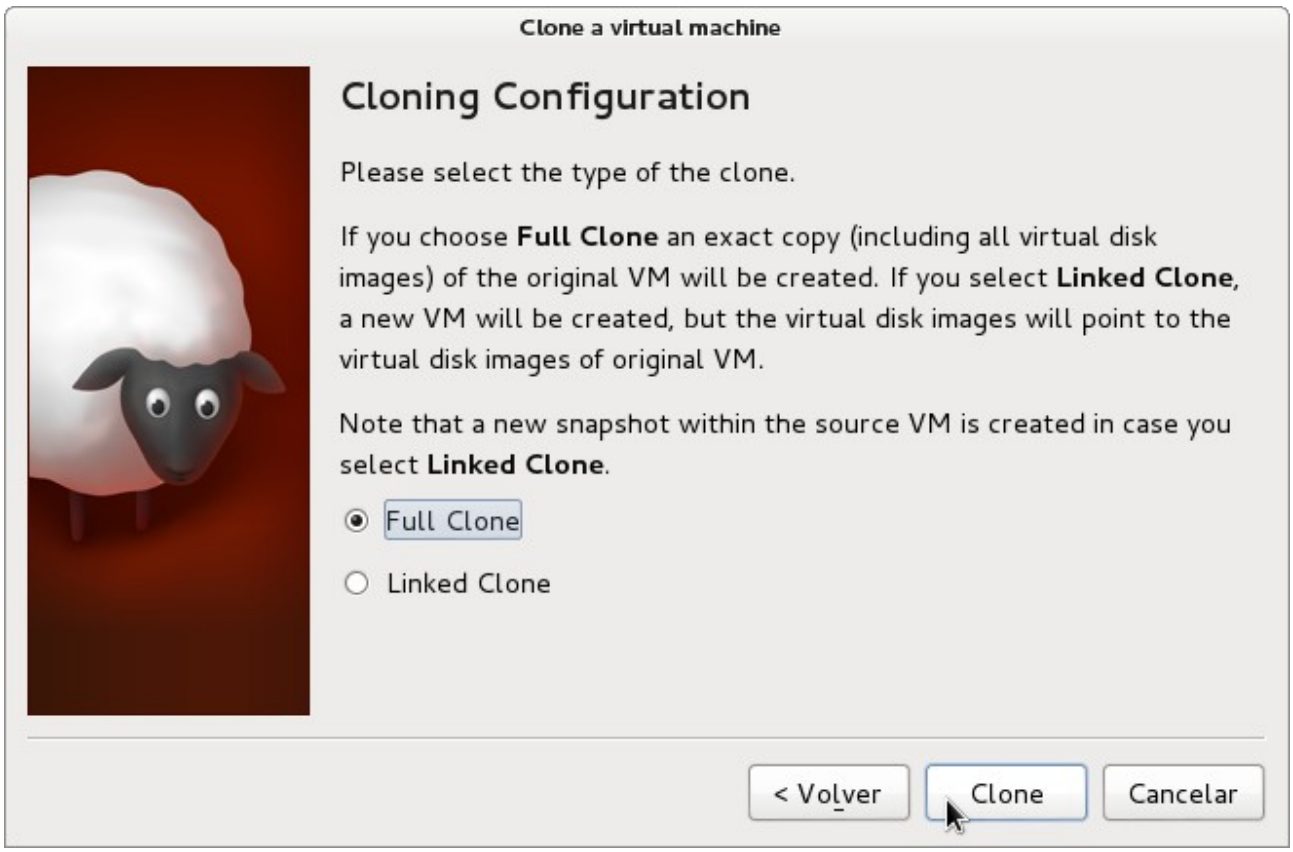
- Verify hostname
  $ hostname

8. Create Hadoop Configuration files

- Move old Cloudera Manager config files
  - $ sudo mv /etc/hadoop/conf /etc/hadoop/conf.alternatives
- Create basic Hadoop config files
  - $ sudo mv /etc/hadoop/conf.empty /etc/hadoop/conf
- Create enviroment vars file
  - $ touch /etc/hadoop/conf/hadoop-env.sh

9. Clone elephant VM

- Shut down elephant on System -> Shut down
- Select elephant on Virtual Box manager and go to menu Machine -> Clone
- Select new names: tiger, horse and monkey.
- Mark reinitialize the MAC address...
- Select Full clone.



10. Start new VM tiger

11. Reconfigure interfaces name
    By default, VirtualBox creates new internal interfaces. We can change the names using these commands:

    - $ sudo nano /etc/udev/rules.d/70-persistent-net.rules
    - Comment with # eth0, eth1 and eth2
    - Rename eth3 -> eth1 and eth4 -> eth2

```
# PCI device 0x8086:0x100e (e1000
SUBSYSTEM=="net", ACTION=="add", DRIVERS=="?*",
ATTR{address}=="08:00:27:6a:1b:86", ATTR{type}=="1", KERNEL=="eth*", NAME="eth1"

# PCI device 0x8086:0x100e (e1000)
SUBSYSTEM=="net", ACTION=="add", DRIVERS=="?*",
ATTR{address}=="08:00:27:75:b5:92", ATTR{type}=="1", KERNEL=="eth*", NAME="eth2"
```

```
# This file was automatically generated by the /lib/udev/write_net_rules
# program, run by the persistent-net-generator.rules rules file.
#
# You can modify it, as long as you keep each rule on a single
# line, and change only the value of the NAME= key.

# PCI device 0x8086:0x100e (e1000)
#SUBSYSTEM=="net", ACTION=="add", DRIVERS=="?*", ATTR{address}=="08:00:27:eb:65:c5", ATTR{type}=="1", KERNEL=="eth*", NAME="e
th0"

# PCI device 0x8086:0x100e (e1000)
#SUBSYSTEM=="net", ACTION=="add", DRIVERS=="?*", ATTR{address}=="08:00:27:9e:e6:60", ATTR{type}=="1", KERNEL=="eth*", NAME="e
th1"

# PCI device 0x8086:0x100e (e1000)
#SUBSYSTEM=="net", ACTION=="add", DRIVERS=="?*", ATTR{address}=="08:00:27:ba:08:a2", ATTR{type}=="1", KERNEL=="eth*", NAME="e
th2"

# PCI device 0x8086:0x100e (e1000)
SUBSYSTEM=="net", ACTION=="add", DRIVERS=="?*", ATTR{address}=="08:00:27:6a:1b:86", ATTR{type}=="1", KERNEL=="eth*", NAME="et
h1"

# PCI device 0x8086:0x100e (e1000)
SUBSYSTEM=="net", ACTION=="add", DRIVERS=="?*", ATTR{address}=="08:00:27:75:b5:92", ATTR{type}=="1", KERNEL=="eth*", NAME="et
h2"
~
~
~
```

12. Configure tiger Network

- Configure IP
  $ sudo nano /etc/sysconfig/network-scripts/ifcfg-eth2
  ...
  IPADDR=192.168.1.2
  ...

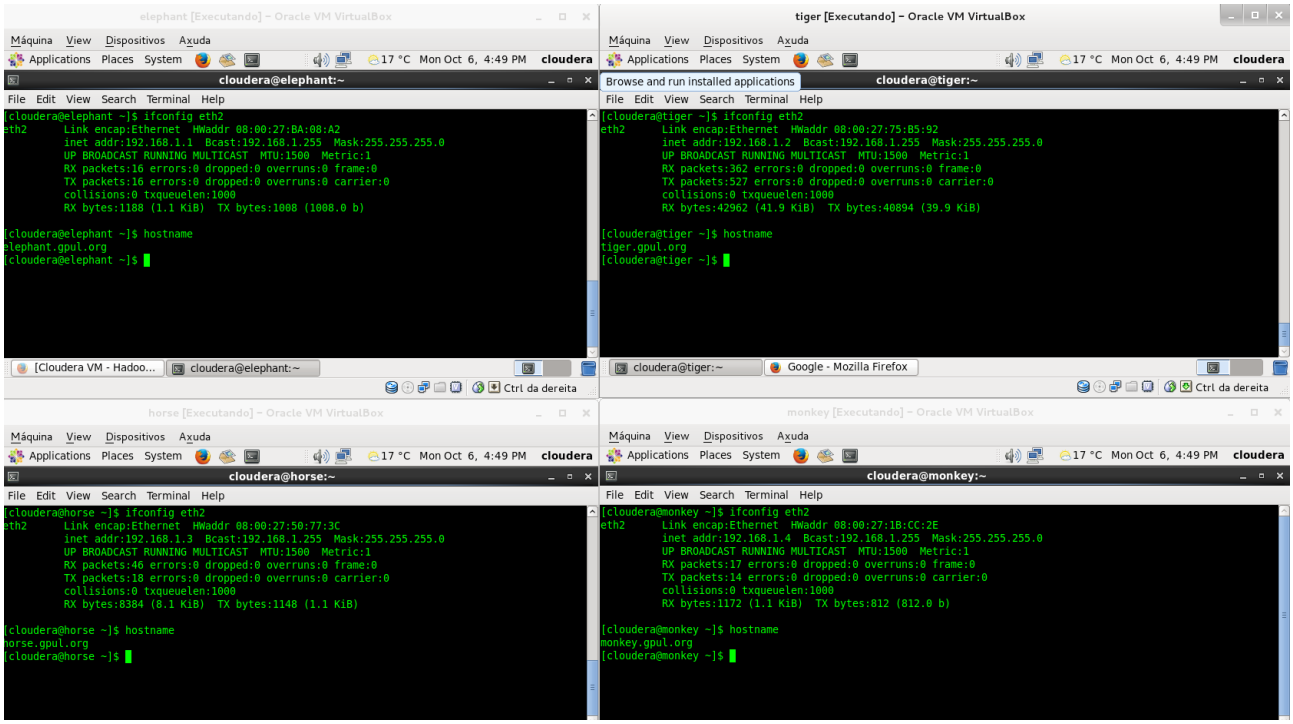
- Configure hostname
  $ sudo nano /etc/sysconfig/network
  NETWORKING=yes
  HOSTNAME=**tiger**.gpul.org

- We need reboot VM after change ethX names
  $ sudo reboot

13. Copy and configure ***horse*** same as tiger in steps 10 to 14
14. Copy and configure ***monkey*** same as tiger in steps 10 to 14

Test ping to elephant, tiger, horse and monkey:
$ ping elephant
$ ping tiger
$ ping horse
$ ping monkey

There should be the cluster prepared like in this picture:


elephant [Executando] – Oracle VM VirtualBox
Máquina View Dispositivos Axuda
Applications Places System
17 °C Mon Oct 6, 4:49 PM cloudera
cloudera@elephant:~
File Edit View Search Terminal Help
[cloudera@elephant ~]$ ifconfig eth2
eth2 Link encap:Ethernet HWaddr 08:00:27:BA:08:A2
inet addr:192.168.1.1 Bcast:192.168.1.255 Mask:255.255.255.0
UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1
RX packets:16 errors:0 dropped:0 overruns:0 frame:0
TX packets:16 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:1000
RX bytes:1188 (1.1 KiB) TX bytes:1008 (1008.0 b)
[cloudera@elephant ~]$ hostname
elephant.gpul.org
[cloudera@elephant ~]$
[Cloudera VM - Hadoo...
cloudera@elephant:~
Ctrl da dereita
tiger [Executando] – Oracle VM VirtualBox
Máquina View Dispositivos Axuda
Applications Places System
17 °C Mon Oct 6, 4:49 PM cloudera
Browse and run installed applications
cloudera@tiger:~
File Edit View Search Terminal Help
[cloudera@tiger ~]$ ifconfig eth2
eth2 Link encap:Ethernet HWaddr 08:00:27:75:B5:92
inet addr:192.168.1.2 Bcast:192.168.1.255 Mask:255.255.255.0
UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1
RX packets:362 errors:0 dropped:0 overruns:0 frame:0
TX packets:527 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:1000
RX bytes:42962 (41.9 KiB) TX bytes:40894 (39.9 KiB)
[cloudera@tiger ~]$ hostname
tiger.gpul.org
[cloudera@tiger ~]$
cloudera@tiger:~
Google - Mozilla Firefox
Ctrl da dereita
horse [Executando] – Oracle VM VirtualBox
Máquina View Dispositivos Axuda
Applications Places System
17 °C Mon Oct 6, 4:49 PM cloudera
cloudera@horse:~
File Edit View Search Terminal Help
[cloudera@horse ~]$ ifconfig eth2
eth2 Link encap:Ethernet HWaddr 08:00:27:50:77:3C
inet addr:192.168.1.3 Bcast:192.168.1.255 Mask:255.255.255.0
UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1
RX packets:46 errors:0 dropped:0 overruns:0 frame:0
TX packets:18 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:1000
RX bytes:8384 (8.1 KiB) TX bytes:1148 (1.1 KiB)
[cloudera@horse ~]$ hostname
horse.gpul.org
[cloudera@horse ~]$
monkey [Executando] – Oracle VM VirtualBox
Máquina View Dispositivos Axuda
Applications Places System
17 °C Mon Oct 6, 4:49 PM cloudera
cloudera@monkey:~
File Edit View Search Terminal Help
[cloudera@monkey ~]$ ifconfig eth2
eth2 Link encap:Ethernet HWaddr 08:00:27:1B:CC:2E
inet addr:192.168.1.4 Bcast:192.168.1.255 Mask:255.255.255.0
UP BROADCAST RUNNING MULTICAST MTU:1500 Metric:1
RX packets:17 errors:0 dropped:0 overruns:0 frame:0
TX packets:14 errors:0 dropped:0 overruns:0 carrier:0
collisions:0 txqueuelen:1000
RX bytes:1172 (1.1 KiB) TX bytes:812 (812.0 b)
[cloudera@monkey ~]$ hostname
monkey.gpul.org
[cloudera@monkey ~]$