



## Apéndice A

# Instalación de un Cluster Rocks

Jorge Zuluaga – Instituto de Física – U. de A.

Última actualización: 5 de diciembre de 2006

Se presenta en este apéndice una guía práctica a la instalación completa de un Cluster Rocks y su configuración básica para realizar tareas de procesamiento distribuido y en paralelo. El proceso de instalación descrito en este apéndice supone una versión de Rocks posterior a la 4.2.1 y la disposición de un Cluster de máquinas de 32 bits con interfaces de red ethernet. Para versiones anteriores de Rocks el procedimiento podría diferir en algunos apartes (sobre todo en la parte de configuración) pero a grandes rasgos sigue las mismas líneas. La guía presentada aquí no cubre la instalación de Rocks en máquinas de 64 bits, Clusters heterogéneos (32/64) o en Clusters con redes que usen una tecnología diferente a Ethernet, aunque una buena parte de los procedimientos descritos se aplica también en estos casos.

Para una guía más completa de la instalación y la configuración se refiere al lector a la guía del usuario de Rocks (<http://www.rocksclusters.org>).

### 1.Requerimientos y prerequisites

Antes de proceder con la instalación se deben satisfacer los siguientes requisitos de hardware:

1. Un conjunto de máquinas de arquitectura similar (compute nodes), cada una con una interfaz de red, disco duro con capacidad para más de 7 GB y memoria RAM superior a 256 MB.
2. Un data switch (o varios) con un número de puertos mayor al doble del número de máquinas disponibles (para darle escalabilidad).
3. Una máquina con 2 interfaces de red, capacidad en disco duro igual o superior a 20 GB, y memoria RAM superior o igual a 512 MB (frontend).
4. Cables de red en número y longitud suficiente.

Se recomienda además tener a disposición:

5. Mueble o Rack con espacio apropiado para los chasis de las máquinas y eventualmente para el frontend, con acceso apropiado a la parte de atrás de los equipos.
6. Una habitación con ventilación o refrigeración adecuada para los niveles de disipación de calor de todos los equipos combinados.
7. Una UPS para alimentar al menos una máquina (el frontend por ejemplo) por más de 10 minutos

Para la instalación del sistema operativo se debe disponer de los siguientes medios:

- Kernel Roll
- Core Roll
- OS Roll, disk 1
- OS Roll disk 2
- Cualquier otro Roll que considere necesario (Condor, Bio, Viz, etc.)

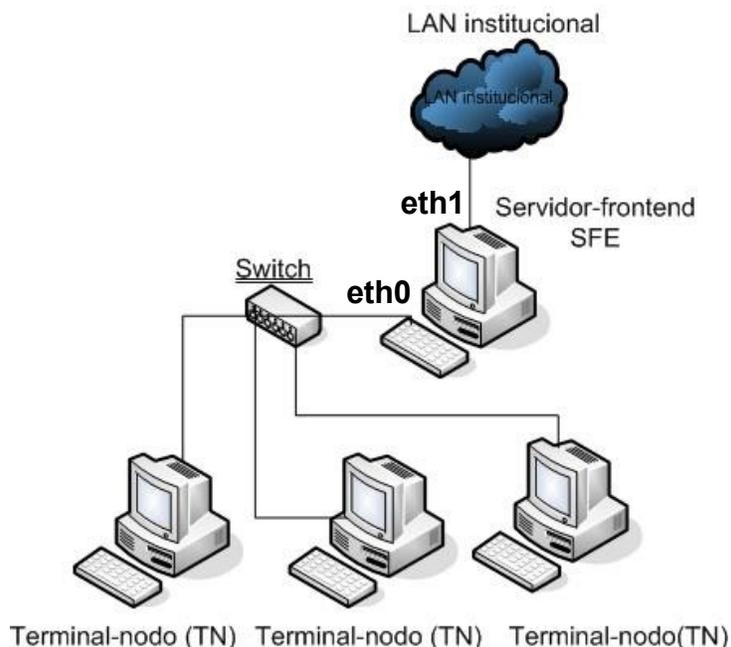
Los medios pueden ser descargados como imágenes de CD o una imagen única de DVD del sitio de descargas de Rocks ([http://www.rocksclusters.org/wordpress/?page\\_id=3](http://www.rocksclusters.org/wordpress/?page_id=3)).

### Tips útiles

Antes de proceder con la instalación se recomienda disponer de información detallada de los equipos que servirán como nodos y del frontend. De particular interés es conocer el fabricante y modelo de las interfaces de red, el número de procesadores efectivos por máquina (1 procesador con HT = 2 procesadores efectivos) y el espacio en disco duro.

## 2.Instalación del frontend

Antes de proceder con la instalación del frontend es necesario asegurarse que las conexiones de la red externa y la red interna del cluster se hagan a la interfaz de red correcta. Rocks asume que la interfaz identificada como 'eth1' por el kernel será aquella que esta conectada a la red externa y la 'eth0' a la red privada del cluster (ver figura).



La manera como el kernel asigna los identificadores eth0 y eth1 es relativamente compleja y varía con el fabricante y modelo de las tarjetas, la board y los puertos a los que están conectadas y la

versión del kernel de Linux (para una discusión del asunto vea <http://linux.dell.com/files/whitepapers/nic-enum-whitepaper-v2.pdf>). Normalmente en Linux es posible cambiar después de la instalación la manera como el sistema asigna los identificadores a las tarjetas de acuerdo a su MacAddress. Sin embargo cuando se instala Rocks es importante que el orden de ellas sea el que se espera de acuerdo al esquema de conectividad del frontend.

El síntoma de que existe un problema en el orden de identificación de las tarjetas de red aparece cuando en los primeros momentos de la instalación el proceso no logra obtener una dirección IP de la red a la que esta conectada la interfaz eth1. En este caso podrían darse 2 situaciones:

1. La red externa tiene servicio de DHCP. En este escenario el hecho de que el proceso de adquisición de una IP por parte del servidor no funcione es síntoma de que el orden de identificación no fue el correcto y la interfaz que va a la red externa seguramente esta identificada como eth0. La solución en este caso puede ser simplemente cambiar intercambiar los cables de conexión y re intentar la adquisición de una dirección IP dinámica. En caso de que las tarjetas para la red externa y privada sean muy distintas (por ejemplo la red privada puede manejarse mejor si se usa una interfaz rápida de 1 a 10 Gbits), la solución más simple puede ser simplemente intercambiar las ranuras a las que están conectadas las tarjetas. Si una de ellas esta integrada a la Motherboard el problema se complica aún más y se sugiere en este caso buscar asesoría especializada.
2. La red externa no tiene servicio DHCP. En este escenario no es posible identificar de forma inmediata cuál identificador se asigna a las interfaces. Para resolver el inconveniente podría ser posible usar un disco de arranque de Linux (Linux rescue, Linux Live CD) que use un kernel de versión no muy distinta a la que usa Rocks (4.2.1 incluye un kernel 2.6) e investigar el modo en el que los identificadores de las NIC son asignados.

### Tips útiles

Para saber el identificador que fue asignado a una NIC específica puede primero consultar la Mac Address (o HW Address) de la NIC asociada a un identificador usando:

#### Comandos 1,1:

```
# ifconfig eth0 | head -n1  
eth0      Link encap:Ethernet  HWaddr 00:0A:5E:4C:94:1A
```

Y luego (en sistemas basados en Redhat) consultar el archivo /etc/sysconfig/hwconf para saber a que NIC (por fabricante y modelo corresponde esta Mac Address:

#### Comandos 1,2:

```
# grep -B 5 -A 5 00:0A:5E:4C:94:1A /etc/sysconfig/hwconf  
bus: PCI  
detached: 0  
device: eth0  
driver: 3c59x  
desc: "3Com Corporation 3c905C-TX/TX-M [Tornado]"  
network.hwaddr: 00:0A:5E:4C:94:1A  
vendorId: 10b7  
deviceId: 9200  
subVendorId: 10b7  
subDeviceId: 1000
```

Una vez resuelto el inconveniente (de existir en primera instancia) de los identificadores de las NICs puede proceder con la instalación. A continuación se describe rápidamente el proceso (que puede ser seguido en detalle en la Rocks User's Guide).

1. Arranque del frontend usando el disco con el Kernel Roll. En el prompt del 'boot' escriba la palabra 'frontend'.

### Tips útiles

Si el proceso de instalación no arranca exitosamente o se detiene antes de avanzar a una etapa posterior consulte al equipo de soporte de Rocks el origen de su problema. Algunos problemas previamente identificados y resueltos por nuestro equipo incluyen:

- Equipos con memoria RAM insuficiente (inferior a 512 MB). Solución: aumentar la cantidad de memoria RAM disponible aunque sea temporalmente durante la instalación.
- Equipos con tarjetas de video no reconocidas propiamente por Rocks. Solución: intentar otras versión del sistema o ponerse en contacto con el equipo de desarrollo de la distribución.

### 2. Pantallazo 1:

- a) Para instalaciones desde medios seleccione en el pantallazo inicial la opción: **CD/DVD-based Roll**. El uso de esta opción en este punto le mostrará una lista de los Rolls contenidos en el disco actualmente presente en la unidad (en el caso de instalación desde un CD solamente aparecerá el Roll kernel). Seleccione el Roll y presione 'submit'
- b) Para insertar nuevos Rolls a la instalación presione siempre en el mismo orden CD/DVD-based Roll (lo que producirá la expulsión cada vez del medio en la unidad), continúe (para que muestre la lista de Rolls en el medio), seleccione los Rolls a instalar, presione submit para incluirlos en la lista. El orden en el que se ingresan los Roll es arbitrarios.
- c) Una vez concluya la inserción de los Rolls continúe con la instalación presionando Next.

### 3. Pantallazo 2: A continuación complete el formulario con la información básica del cluster. Tenga en cuenta las siguientes instrucciones y recomendaciones:

- a) Fully Qualified Host Name (FQHN): use aquí el nombre completo del frontend tal y como es reconocido en la red externa. Consulte con su administrador de redes el nombre que su máquina recibirá en esa red. Este parámetro es importante por qué será utilizado en muchos aspectos de la configuración del cluster y sus herramientas. Si en la red a la que esta conectado su equipo no existe un esquema de asignación de nombres entonces use uno que obedezca a una regla lógica de asignación, nombre.dominio
- b) Cluster Name: use aquí el nombre completo de su herramienta. Puede usar mayúsculas y minúsculas cómo "Cluster Escuela de Ingeniería". El nombre completo será usado para identificar el cluster en piezas de software como Ganglia y otras.
- c) Certificate organization (organización a la que pertenece el cluster), Certificate Locality (ciudad), Certificate state (estado o departamento en el que esta la ciudad), Certificate country (abreviatura del país – ver una lista completa en <http://www.bris.ac.uk/Support/Network/ipcodes1.html>)
- d) Contact: e-mail de contacto (use uno válido)
- e) URL: dirección de la pagina del cluster o de la organización a la que pertenece.

f) Latitude/Longitud: latitud y longitud geográfica de la locación en la que se encuentra el cluster en el format N|S###.### E|W####.### (ejemplo para Medellín Colombia N6.2 W75.5)

### Tips útiles

El instalador de Rocks usa por defecto una configuración de teclado en inglés. Para teclados latinoamericanos o en español tenga presente que el símbolo arroba se encuentra con la combinación de teclas SHIFT+2

4. **Pantallazo 3:** A continuación se configura los parámetros básicos de la red privada. El único detalle de interés para tener presente en este punto es el de si la red externa esta configurada también usando IPs de clase A (10.\*.\*). En ese caso las Ips de la red privada del cluster y de la red externa colisionan produciendo una incompatibilidad. Se recomienda usar una asignación de IPs distinta para la red del cluster. Así por ejemplo:

IP: 192.168.1.1  
Netmask: 255.255.0.0

En este caso las IPs de los nodos se asignarán de la forma 192.168.255.254, 192.168.255.253, 192.168.255.252, en lugar de 10.255.255.254, como se explico en la guía.

5. **Pantallazo 4:** La configuración de los parámetros de red para la interfaz eth1 no se debe modificar en caso de que la red externa tenga DHCP. En esta situación los valores de los parámetros habrán sido obtenidos directamente del servicio DHCP en la red.

6. **Pantallazo 5:** Se fija a continuación el password de Root. para usuarios con un teclado en español o latinoamericano se recomienda usar el password de root más sencillo posible (no usar caracteres especiales) para evitar inconvenientes cuando se configure apropiadamente el teclado en la fase de post instalación.

7. **Pantallazo 6:** configuración del huso horario y del servidor de protocolo de tiempo (NTP). En redes protegidas con un firewall es común que los puertos con los que trabaja el NTP estén protegidos de modo que el servicio de cualquier modo deberá detenerse después de instalado el frontend.

8. **Pantallazo 7:** selección del mecanismo de particionado. Para clusters dedicados se recomienda utilizar el esquema de auto particionado. El particionado por defecto que crea Rocks realiza el siguiente esquema de particiones y puntos de montaje:

<i>Partición</i>	<i>Punto de montaje</i>	<i>Tamaño por defecto</i>	<i>Tipo</i>
a1*	/	7.7 GB	Ext3
a2	/var	3.9 GB	Ext3
a3	-	1 GB	Swap
a4	/state/partition1	Grow	Ext3

- Vale por /dev/hda1 o /dev/sda1

Es posible cambiar el esquema de particionado si se desean incluir en el frontend particiones

especiales para respaldos, software, sistemas de archivos distribuidos, etc. En ese caso se debe sin embargo tratar de respetar las elecciones hechas arriba para el punto de montaje de los sistemas de archivos que usa por defecto rocks.

En el caso de reinstalación del frontend y si no quiere formatearse alguna de las particiones se debe elegir la opción de particionado manual.

9. Una vez completo el proceso de configuración comienza propiamente el proceso de instalación de la distribución. Esta fase incluye (puede incluir) las siguientes tareas para la persona que realiza la instalación:

a) Personalización del esquema de particionado (Si se eligió la opción de particionado manual)

b) La persona debe proveer los medios que contienen los Rolls que fueron seleccionados para la instalación. El instalador los irá pidiendo uno a uno esos medios (si se trata de Cds). La información con los paquetes y los paquetes mismos son transferidos al disco duro antes de instalarse.

10. Una vez completo el proceso anterior el proceso comienza a instalar todos los paquetes contenidos en los Rolls y finaliza con la ejecución de todos los scripts de post instalación definidos con las herramientas incluidas en la distribución.

### 3. Configuración del frontend

Una vez instalado completamente el frontend se puede proceder a ajustar algunos detalles de la configuración de la máquina y la preparación de la distribución que será instalada en los compute nodes.

Para usuarios con teclados latinoamericano o español es muy conveniente cargar el mapa de teclado correspondiente. Para hacerlo se usa el comando:

Comandos 1.3:

```
# loadkeys /lib/kbd/keymaps/i386/qwerty/i386/es.map.gz
```

Para garantizar que la configuración del teclado se mantenga aún después de reiniciado el frontend se puede agregar esta línea al archivo rc.local:

Comandos 1.4:

```
# echo loadkeys /lib/kbd/keymaps/i386/qwerty/i386/es.map.gz >> /etc/rc.local
```

En lo sucesivo puede ser cómodo contar con una consola gráfica en las tareas de instalación y configuración del cluster. Para ello se puede ejecutar la siguiente secuencia de comandos:

Comandos 1.5:

```
# service gpm start  
# service xfs restart  
# system-config-display
```

**chkconfig --levels 35 gpm on**

chkconfig –levels 35 xfs on

Este último aplicativo permite ajustar las propiedades del modo gráfico (resolución profundidad de color, aceleración de hardware, entre otras).

Para iniciar el modo gráfico se puede usar:

Comandos 1,6:

```
# startx
```

Rocks utiliza por defecto Gnome como gestor de escritorio. La configuración de teclado de Gnome normalmente difiere de aquella que viene con el sistema X11 (y que se hereda de la configuración de teclado en la consola de texto). Al iniciar Gnome por primera vez es necesario entonces configurar manualmente el mapa del teclado. Esto se consigue usando la herramienta de configuración “keyboard” que se encuentra en el submenú “System settings” del menú “Application”.

Para que la configuración permanezca se sugiere reiniciar limpiamente el modo gráfico (usar el botón “Actions>Log out”) y reiniciarlo nuevamente. Una vez suceda esto el GNOME preguntará por cuál debe ser la configuración que seguirá usando por defecto para los dispositivos a lo que se deberá responder “X11 settings”.

Una vez habilitada una consola gráfica es posible realizar con más comodidad algunas de las tareas de configuración que se describen a continuación.

Para continuar con la configuración del frontend, antes de la instalación de los nodos de computo se recomienda completar las siguientes tareas:

1. Creación de cuentas de usuario. Utilice las instrucciones proveídas en la guía 3, sección 3.3 para crear las cuentas de usuario en el frontend.
2. Preparación del /export/apps. Cree la estructura de subdirectorios en /export/apps tal y como se describe en la guía 4, sección 4.1, comandos 4,17.
3. Instalación de software desde las fuentes (tarballs). Instale todo el software adicional en el frontend que desea instalar en el cluster desde tarballs sobre el directorio /export/apps. Ver detalles en guía 4, sección 4.1.
4. Instalación de software desde rpms. Consiga e instale el software no presente en la distribución y de cuyos binarios en RPM se pueda disponer.
5. Preparación y creación de la distribución. Una vez se garantiza que todos los paquetes en RPM conseguidos pueden ser instalados en el frontend y antes de crear la distribución para los nodos, configure el sistema de instalación para que incluya esos mismos paquetes (ver instrucciones detalladas del proceso en la guía 4, sección 4.1).
6. Finalmente configure el sistema 411 para que sincronice en todos los nodos además de los

archivos de configuración por defecto (/etc/passwd, /etc/group, etc.) otros archivos que pueda ser de utilidad compartir a través del cluster (p.e. /etc/profile, /etc/bashrc, /etc/ld.so.conf, etc.) Para ello refierase a la guía 3, sección 3.2.

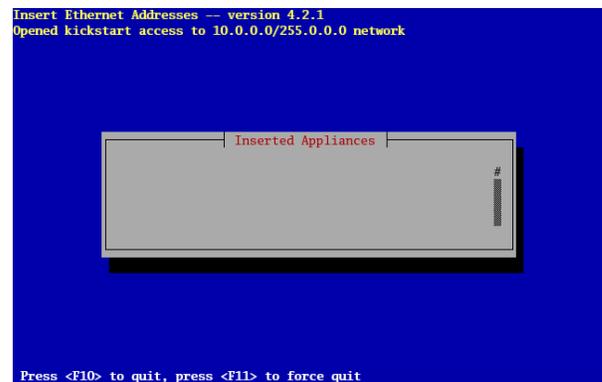
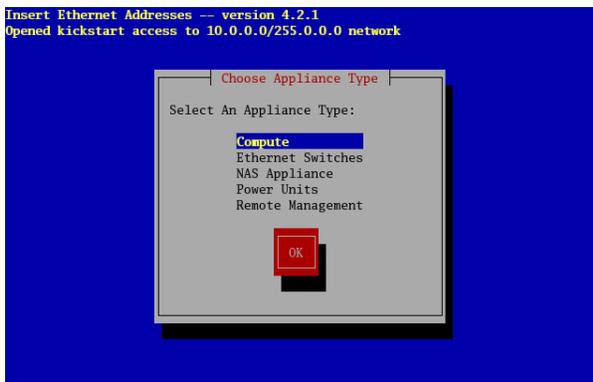
Una vez se ha completado el procedimiento anterior comience la instalación de los nodos.

## 4.Instalación de los nodos

Para comenzar con la instalación de los nodos se debe iniciar 'insert-ethers' un aplicativo que permite la asignación automática de las IPs a los nodos y la creación de la base de datos de dispositivos conectados al Cluster. Para iniciar insert-ethers ejecute:

Comandos 1,7:

```
# insert-ethers
```



Si no se cuenta con un switch manejado seleccione directamente 'Compute' para comenzar el proceso de detección de los nodos conectados a la red privada del cluster.

Una vez iniciado insert-ethers comience a arrancar secuencialmente los nodos de computo, o bien usando para ello el disco del Kernel Roll o usando PXE en caso de estar disponible y habilitado en el equipo.

### Tips útiles

Para instalaciones usando medios (CD o DVD) se recomienda crear varias copias del CD del Kernel Roll para agilizar el proceso de instalación de los nodos. Tenga presente que si instalo el frontend usando DVD pero los nodos no tienen unidad lectora de este tipo de medios puede descargar solamente la imagen del Kernel Roll para la instalación de los nodos.

Si desea que sus nodos queden numerados siguiendo un orden riguroso relacionado con la manera como se disponen físicamente no se apresure con la instalación simultánea de muchos de esos nodos. La asignación de la IP y el nombre de cada nodo se irá haciendo, la primera vez, en el orden en el que vayan llegando los requerimientos de IP al frontend. Una vez registrada la Mac Address del equipo su IP y nombre quedan determinados definitivamente. Evite conectar equipos en la red privada (portátiles, otros Pcs, etc.) que no vayan finalmente a formar parte del Cluster. Insert-ethers asignará a esos equipos una IP y un nombre y su eliminación de la lista de

nodos una vez no estén conectados a la red no es una tarea sencilla.

### Tips útiles

Se puede chequear el avance en la instalación de uno de los nodos usando el comando:

#### Comandos 1,8:

```
# rocks-console compute-0-0
```

rocks-console funciona solamente si se tiene acceso a una consola gráfica en el frontend.

## 5. Configuración final de los nodos

Si la instalación de los nodos es exitosa entonces no hace falta mayor cosa para comenzar a utilizar el Cluster. Por precaución se pueden realizar algunas pruebas sencillas para verificar que todo este funcionando normalmente en la plataforma.

1. ¿El comando cluster-fork funciona correctamente?

#### Comandos 1,9:

```
# cluster-fork w
```

No:

- Los nodos piden password de root. Ocurrió algún problema con la creación de la clave pública y privada del root.
- Ningún nodo responde. El proceso de instalación de los nodos no se completo exitosamente.

2. Examine la salida de 'qstat -f'. Si la salida no es normal o muestra algunos nodos en estado 'au':

#### Comandos 1,10:

```
# qstat -f
```

- Revise que los nodos estén funcionando

3. ¿Un usuario puede conectarse exitosamente en el frontend?

#### Comandos 1,11:

```
# ssh fulano@cluster.dominio
```

No:

- Revise que el usuario si exista
- El usuario no encuentra su home directory: revise el archivo de configuración /etc/auto.home y recargue el servicio de autofs (Comandos 3,7)
- Después del reload el usuario todavia no encuentra su home directory: apague los nodos y reinicie el frontend. Después de reiniciar el frontend, encienda los nodos.

4. Una vez adentro del frontend, ¿puede el usuario conectarse transparentemente a otros nodos?

No:

- Al usuario se le pide password y una vez ingresa no se encuentra su home directory: re sincronice los archivos de configuración en todo el cluster (comandos 3,13 o 3,14) y reinicie en todo el cluster el servicio autofs.