Instalación y Configuración de una Sala Cluster como una Solución de Bajo Costo para el Acceso a Tecnología de Computación Distribuida en Latinoamérica

Installation and Configuration of a Cluster-Room as a Low Cost Solution for the Access to Distributed Computing Technologies in Latin America

Jorge I. Zuluaga Callejas
Grupo de Física y Astrofísica Computacional, FACom, Instituto de Física, Universidad de Antioquia, Colombia
jzuluaga@fisica.udea.edu.co

Alvaro E. Ospina Sanjuan
Grupo de Microelectrónica, Escuela de Ingenierías, Universidad Pontificia Bolivariana, Colombia
alvaro.ospina@upb.edu.co

Resumen

Se presenta procedimientos para la instalación y configuración de una sala de computo pública (sala de terminales) que pueda realizar además las funciones de un Cluster Beowulf para computación de alto desempeño y otras tareas de computo distribuido (en adelante Sala Cluster). Los procedimientos se basan en la metodología general para la instalación y configuración de salas cluster publicadas previamente por los autores aplicadas en particular para la instalación de salas clusters usando GNU/Linux Rocks y Windows XP como sistema operativo del cluster y sistema operativo utilitario respectivamente. Para instituciones que no cuentan con los recursos necesarios para la adquisición de una herramienta de computación dedicada, los procedimientos descritos aquí permiten el montaje de un cluster de cómputo parcialmente dedicado usando el tiempo de inactividad de recursos de cómputo existentes en salas públicas. Los resultados y experiencias con la aplicación de estos procedimientos en dos instituciones de educación superior en Colombia, así como las dificultades enfrentadas en el proceso y el impacto que las salas han tenido en la investigación y educación en áreas de la computación distribuida se presentan también en este trabajo

Abstract

We present here procedures for the installation and configuration of a public computer room (room of computer terminals) that can also be used as a Beowulf Cluster for high performance computing and other distributed computing tasks (thereafter Cluster-Room). The procedures presented here are based in a general methodology previously published by the authors which is applied in particular for the installation of a cluster room using GNU/Linux Rocks and Windows XP as the cluster operative system and the utilitary operative system respectively. For institutions with limited economical resources for the acquisition of a dedicated distributed computing platform, specially in Latin America, the procedures described here are specially suited for the installation of a computer cluster using the idle time at night and weekends of computer resources in public computer rooms. The results of the application of these procedures in two Universities in Colombia, the challenges faced in the process and the impact that those Cluster-Rooms have had in the research and education activities in distributed computing topics are also presented in this work.

1. Introducción

Se presenta en este trabajo los resultados del desarrollo y aplicación de una metodología que permite el montaje o la reconfiguración de una sala de cómputo pública para que pueda utilizarse como un cluster de cómputo parcialmente dedicado. La metodología fue introducida por primera vez por uno de los autores (J. I. Zuluaga) durante el Encuentro de Investigación sobre Tecnologías de Investigación Aplicada, EITI-2005 [8] y presentada en forma más general por Zuluaga y Ospina en 2006 [7]. Llamaremos en lo sucesivo a una plataforma concebida de esta manera una *Sala Cluster*. La instalación de un cluster parcialmente dedicado con el modelo de Sala Cluster permite el acceso a una plataforma de cómputo

de alto rendimiento (en lo sucesivo *HPC*) utilizando recursos disponibles en salas de cómputo públicas. Una Sala Cluster puede servir también como una *entry level solution* para instituciones o dependencias que no pueden invertir en un recurso dedicado para HPC, pero que necesitan de una herramienta de esta naturaleza como un banco de pruebas de estas tecnologías o para capacitación en el área. El modelo es también apropiado para la instalación de sitios en un Grid inter institucional que potencie la utilización de recursos en tiempo de inactividad.

Diversos esfuerzos se han hecho para llevar plataformas de computación de alto rendimiento de bajo costo a entornos de naturaleza educativa y formativa [12], [6]. Se han desarrollado distribuciones de GNU/Linux que pueden ser desplegadas en minutos usando sistemas de booteo desde CD-ROMs y se las ha utilizado en ambientes educativos [12], [2]. Este tipo de soluciones, si bien permite el acceso casi inmediato en ambientes heterogéneos y poco estructurados a sistemas de recursos de computo distribuidos, tienen la desventaja de tener vidas relativamente cortas e importantes limitaciones en temas como el almacenamiento permanente de resultados de cálculos o la instalación de software no incluido en las distribuciones utilizadas. Otros se valen de recursos reciclados para el montaje de sistemas de cómputo distribuido con aplicaciones principalmente en tareas de investigación formativa o para educación [11]. Este tipo de aproximación tiene también limitaciones reconocidas. La alta heterogeneidad en los recursos que pueden ser reciclados, dificultades en la instalación de versiones recientes del software en arquitecturas o sistemas obsoletos y las limitadas capacidades de los mismos recursos debido a su antigüedad. Sistemas de computación en Grid instalados para explotar el tiempo de inactividad de equipos de escritorio de empleados o profesores pueden ser también utilizados con el fin de proveer acceso barato a esta tecnología. Si bien esta es una solución viable para ofrecer recursos de cómputo casi ilimitados para enormes problemas de cálculo tiene limitaciones para ofrecer ambientes estables para la enseñanza o la investigación formativa en el área. La idea detrás del concepto de Sala Cluster ofrece soluciones a algunas de las problemáticas que estos enfoques exhiben. Primero utiliza recursos preexistentes pero no reciclados. Normalmente los equipos disponibles en salas de cómputo públicas son equipos con configuraciones y prestaciones aceptables. Segundo ofrece un ambiente homogéneo de recursos que imita de cerca el que un estudiante o investigador podría encontrar en plataformas de computo intensivo dedicadas. Tercero si bien se vale del tiempo de inactividad de los recursos lo hace también de manera homogénea, al definirse (según las políticas administrativas) tiempos específicos de uso de la sala como cluster.

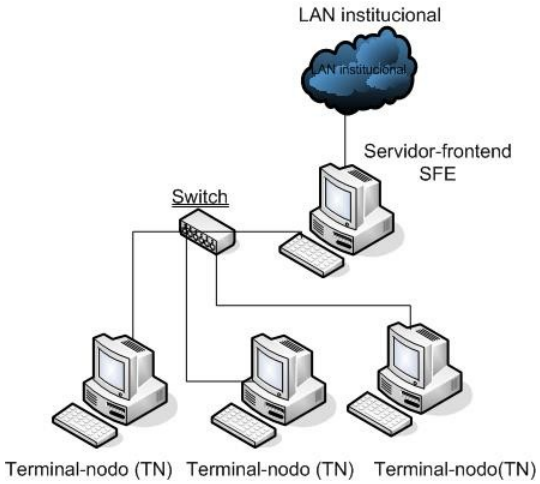

Figura 1. Arquitectura básica de la Sala Cluster como es concebida en [7]

Según la metodología concebida por Zuluaga y Ospina [7], la instalación de la Sala Cluster implica la solución al siguiente conjunto de condicionantes básicos: 1) Instalación y configuración de un sistema de inicio dual (*dual boot*) para dos sistemas operativos: uno dedicado a los servicios propios de la sala (*sistema operativo utilitario*) y otro en el que se aglomere los recursos y herramientas para formar el cluster de cómputo (*sistema operativo del cluster*). 2) Instalación y configuración del sistema operativo GNU/Linux para operar la sala como un cluster. 3) Adaptación o implementación de herramientas y metodologías para la instalación automática y masiva de los sistemas operativos en la sala cluster. 4) Implementación o adaptación de herramientas que faciliten la administración efectiva de la sala cluster. 5) Definición de unas políticas de uso de la plataforma.

En la figura 1 se muestra un diagrama de la arquitectura básica de la Sala Cluster como es concebida en [7]. En ella un Servidor Frontend (en adelante SFE), que tiene 2 interfaces de red, se conecta a la LAN institucional y a un concentrador o swiche que sirve a la Sala. Todos las terminales-nodo (en adelante TN) están conectadas directamente al concentrador o swiche y son enrutadas hacia la red externa a través del SFE.

En este trabajo presentamos los detalles técnicos específicos de la implementación de la metodología propuesta por Zuluaga y Ospina al caso de una Sala Cluster usando como sistema operativo del cluster (en adelante SO del cluster) GNU/Linux Rocks [10], (http://www.rocksclusters.org) y como sistema operativo Utilitario (en adelante SO utilitario) Windows XP. Este trabajo se estructura de la siguiente manera: en la sección 2 se describen las condiciones mínimas para iniciar el proceso de instalación de la

Sala Cluster (condiciones de Hardware, Software y parámetros de configuración). En la sección 3 se describe detalladamente el proceso de instalación de la Sala-Cluster, incluyendo la instalación de GNU/Linux y la replicación sistemática de Windows en las terminales de la sala. La sección 4 describe algunas recomendaciones especiales relacionadas con la configuración del sistema. En la sección 5 se presentan los resultados de la aplicación de estos procedimientos en situaciones específicas y se discuten algunos de los problemas y retos de la administración de una plataforma de esta naturaleza.

2. Preparación

Antes de comenzar la instalación de la Sala Cluster es necesario satisfacer unos requerimientos mínimos de hardware y de software y disponer de información para la configuración de la plataforma.

2.1. Requerimientos de hardware

Las máquinas que servirán como TN de la sala cluster deberán satisfacer las siguientes condiciones mínimas de hardware: 1) Disco duro con capacidad total superior a 40 GB, 2) Más de 256 MB de memoria RAM (preferiblemente 512 MB); este requerimiento depende de la versión de Rocks que se va a instalar; 3) Una interfaz de red, 4) Unidad de CD-ROM o floppy (alternativamente se puede tener habilitado un sistema de inicio por red usando PXE), 5) Mouse y teclado (el mismo juego de caracteres en el teclado simplifica el proceso de configuración.)

El SFE puede tener una configuración de hardware distinta de los TN. En este caso además de las condiciones exigidas para los nodos se deben cumplir los siguientes requisitos adicionales: 1) disco duro con capacidad superior o igual a 80 GB, 2) más de 512 MB reales en RAM (prestar atención a los sistemas donde parte de la RAM se usa para la tarjeta de video), 3) una tarjeta de red adicional para la salida a la red institucional o a Internet. Como requisitos opcionales para el SFE tenemos: 1) Un segundo disco duro o un dispositivo de almacenamiento masivo (tape, (CD)DVD+RW) que servirá entre otras cosas para los respaldos del sistema.

Para la interconectividad entre los equipos de la sala cluster se deberá contar con un swiche o concentrador. Se recomienda el uso de un dispositivo con un backplane con la más alta capacidad posible (1 Gbit/s o más) de modo que los puertos operen a la más alta velocidad de transferencia posible.

2.2. Requerimientos de software

Para la instalación de GNU/Linux Rocks como SO del Cluster se requiere contar con los medios básicos de instalación, Rolls [5], enumerados a continuación: 1) Kernel/Boot Roll, 2) Core Roll (base+hpc+ganglia+grid+area51+webserver), 3) Service Pack Roll, 4) OS disk1 Roll, 5) OS disk2 Roll. Dichos medios pueden ser descargados directamente del sitio Web de Rocks (http://www.rocksclusters.org).

Para el arranque y la instalación de los TN se puede usar el sistema PXE (*Pre-Boot Execution Environment*) que permite el arranque de las TN sin medios físicos (CD-ROM) a través de la red. El sistema debe estar soportado por la BIOS de los equipos. De no estarlo se puede utilizar un sistema de inicio PXE usando floppy.

La instalación del SO utilitario requiere la disposición de medios y licencias de instalación de la versión autorizada para los equipos de la sala. Además de los medios de instalación del SO se requiere la disposición inicial de los medios necesarios para la instalación de las aplicaciones específicas que se incluirán como parte del software utilitario de la sala.

Se recomienda que el SO utilitario disponga como mínimo de aplicaciones que permitan la lectura del sistema de archivos de GNU/Linux desde Windows (p.e. *explore2fs*) y una aplicación cliente/servidor de ssh (p.e. *openSSH*).

Adicionalmente si se desea que las TN sirvan como terminales también bajo GNU/Linux, se recomienda disponer de los medios y/o instaladores de algunas aplicaciones que no están incluidas en los medios de Rocks. Entre estas herramientas se recomienda instalar como mínimo, un editor de textos gráfico (*OpenOffice, abiword,* etc.), un visor de documentos pdf (*Acrobat Reader, xpdf,* etc.), un browser gráfico (*Mozilla, Firefox, Opera,* etc.) y un visor y/o editor de imágenes (*Gimp*).

2.3. Información para la configuración

Antes de proceder con la instalación se recomienda disponer de la siguiente información que será de vital importancia para la configuración de la Sala Cluster:

1) Asignación del identificador de las interfaces de red del SFE. El kernel de GNU/Linux asigna a las interfaces de red un identificador, *eth0, eth1, eth2,* etc., que depende de distintos factores. Antes de proceder es necesario conocer cuál identificador es asignado a cuál interfaz para que ellas estén conectadas apropiadamente a la red institucional y a la red del cluster. Rocks por defecto asume que la interfaz nombrada como *eth0* esta conectada a la red del cluster y la interfaz *eth1* a la red institucional.

2) Información completa sobre los parámetros de la red institucional, dirección IP, *Fully Qualified Host Name* (FQHN) del SFE, puerta de enlace predeterminada y servidores de nombre de dominio (DNS).

3) Dirección de e-mail del administrador del cluster.

4) Ubicación del cluster, latitud y longitud geográfica (esta información es utilizada por algunas herramientas de monitoreo para representar la ubicación en el planeta del cluster).

3. Instalación de la Sala Cluster

La instalación básica de las componentes de software de la Sala Cluster se realiza en 5 grandes etapas tal y como es descrito en detalle en [7]:

1) *Instalación y configuración básica del SFE.* El SFE se instala como normalmente se instala un frontend de un cluster Rocks. Esta máquina debe estar dedicada exclusivamente para cumplir esta función.

2) *Preparación de la distribución del SO del cluster para las TN.* En esta etapa se configura la distribución de Rocks que será instalada en las TN, incluyendo el esquema de particionado del disco y los paquetes adicionales que serán instalados en ellas (si es el caso) para que cumplan las funciones de terminales GNU/Linux.

3) *Instalación de la Terminal Nodo de Prueba (TNP).* Esta componente servirá como banco de pruebas e imagen para la instalación completa del sistema en las demás TN de la sala. En esta etapa se instala y configura el SO utilitario y posteriormente se clona y copia para usarlo como base de la instalación en las demás TN de la Sala Cluster.

4) *Instalación y configuración del SO del cluster en todas las TN.* En esta fase se utilizan los mecanismos automáticos para la instalación de Rocks en los nodos del cluster. Para ello Rocks usa el mismo esquema de particionado y paquetes adicionales que uso en el TNP.

5) *Replicación de la Instalación del SO utilitario en las TN.* Una vez instalado el SO del Cluster en todas las TN se procede a replicar la imagen del SO utilitario que fue preparada, a partir de la instalación que de este último se hizo en el TNP.

A continuación se describen algunos de los detalles más importantes de las etapas de instalación. Detalles específicos (comandos, archivos de configuración, etc.) se presentan en la documentación completa del procedimiento publicada en el sitio Web oficial de la iniciativa http://urania.udea.edu.co/cluster-room.

3.1. Instalación del SFE

La instalación del SFE no reviste inicialmente ninguna complicación, ni es muy diferente a la instalación del frontend de un cluster Rocks dedicado. El procedimiento para GNU/Linux Rocks se describe detalladamente en la documentación de la distribución. Como una recomendación particular se sugiere escoger la opción de particionado manual en lugar del particionado automático que viene con la distribución. Se sugiere incluir en esta parte además de las particiones por defecto (partición swap, raiz, directorios casa, etc.) una partición para respaldos y una para almacenar la imagen del SO utilitario que se usará mas adelante. En la tabla 1 se presenta un esquema de particionado apropiado para un SFE con un disco duro SATA de 80 GB y una RAM de 1 GB (se usará la notación de Rocks para los puntos de montaje de las particiones. Los valores indicados son aproximados):

Tabla 1. Ejemplo de esquema de particionado sugerido para un SFE con un disco de 80 GB y RAM de 1GB.

Part.	Punto de montaje	Tamaño aproximado	Tipo de FS
sda1	/	8 GB	Ext3
sda2	N/A	2 GB	Swap
sda3	/var	5 GB	Ext3
sda5	/state/partition1	25 GB	Ext3
sda6	/backup	20 GB	Ext3
sda7	/images	20 GB	Ext3

3.2. Configuración del SFE

Una vez completa la instalación del SFE se deben ajustar un par de detalles sobre la configuración del sistema antes de comenzar con las siguientes etapas de la instalación. En particular para facilitar las tareas sucesivas y el monitoreo posterior de la Sala Cluster la consola del SFE deberá estar configurada apropiadamente para reconocer el teclado en su idioma, el mouse y ejecutar el sistema de ventanas. Rocks con el Roll Viz configura automáticamente el sistema de ventanas. El teclado sin embargo deberá configurarse manualmente para un juego de teclas Latinoamericano o en Español (ver documentación de la sala cluster o documentación del comando loadkeys.)

La configuración de algunos parámetros de red, no ajustados en la instalación, puede realizarse en esta etapa. En particular dado que las TN estarán conectadas a la red externa de la sala enrutadas por el SFE, deberá configurarse de forma apropiada el servicio de nombres (DNS) y algunos aliases a máquinas en la red externa que puedan ser accedidas muy frecuentemente por los usuarios.

3.3. Preparación de la distribución del SO del cluster para las TN

La instalación del SO del cluster en las TN debe ser un proceso rápido y que requiera una mínima intervención del administrador. Esto garantizará el despliegue rápido de la plataforma, su automatización y las tareas de actualización del SO y de las aplicaciones instaladas en él. Rocks utiliza para ello un sistema que se basa en el mecanismo de *Kickstart* de RedHat, además posee otros mecanismos específicos desarrollados para la distribución [9].

En el sistema de instalación del cluster que usa Rocks, antes de proceder con la instalación de los nodos, debe proveerse la información completa sobre las características de la distribución que será instalada en esas máquinas. Rocks pre-configura esta información basándose en la instalación realizada en el SFE. Los archivos necesarios para configurar la distribución que será instalada en los nodos se encuentran en el directorio */export/home/install/site-profiles/<version>/nodes* (en adelante denominado *directorio de configuración personalizada de la distribución*, CPD). En nuestro caso, sin embargo es necesario modificar algunas de las características de la distribución de los nodos para garantizar que se comporten como TN de la sala cluster. Entre las cosas que deben de configurarse de forma especial se encuentran:

1) *Información de los paquetes adicionales para las TN.* A la distribución por defecto que prepara Rocks con paquetes básicos para la operación de los nodos, deberá agregarse por un lado la lista de los paquetes que sean necesarios para la operación de los nodos como terminales del Cluster (sistema de ventanas X, gestor de escritorio, compiladores y librerías gráficas, etc.) y de otro los paquetes y utilidades adicionales. En Rocks la lista de paquetes y utilidades adicionales es agregada en el archivo *extend-compute.xml* ubicado en el directorio CPD, cuya sintaxis es descrita en detalle en la documentación de Rocks.

2) *Información sobre el particionado de las TN*. La información sobre la manera como el instalador deberá particionar y formatear el disco duro de los nodos es uno de los aspectos críticos de esta etapa de preparación. El aspecto más importante para tener presente aquí es la preparación de la partición sobre la que se instalará el SO utilitario, que por ejemplo en el caso de Windows impone algunas restricciones importantes. En Rocks el esquema de particionado se configura en el archivo *replace-autopartition.xml* ubicado en el directorio CPD. Como un ejemplo se presenta en la tabla 2 el esquema de particionado sugerido para un TN con disco SATA de 40 GB y RAM de 1 GB.

Tabla 2. Ejemplo del esquema de particionado sugerido para las TN suponiendo un disco de 40 GB y RAM de 1GB.

Part.	Punto de montaje	Tamaño aproximado	Tipo de FS
sda1	/utilos	18 GB	VFAT
sda2	/	8 GB	Ext3
sda3	N/A	2 GB	Swap
sda5	/state/partition1	12 GB	Ext3

Debe tenerse presente que el tamaño disponible para la partición utilizada para el SO utilitario (montada aquí sobre */utilos*) debe ser adecuado para la instalación de todas las aplicaciones adicionales necesarias para la sala. Así mismo la partición */state/partition1* deberá tener capacidad suficiente para el copiado sobre ella de la imagen del SO utilitario. Este es un aspecto crítico en la instalación de la sala cluster. El esquema de particionado ilustrado en la tabla 2 se debe codificar en el archivo de configuración *replace-autopartition.xml* que se encuentra en el directorio CPD, siguiendo los lineamientos ofrecidos en la documentación de Rocks. Es importante anotar que la partición utilizada para el SO utilitario, se configura como VFAT, pero luego se le cambiará al tipo NTFS al momento de instalar Windows XP en los computadores.

Una vez se ha completado satisfactoriamente la configuración de la distribución, en Rocks debe crearse la distribución misma. Esto se consigue usando el comando *rocks-dist* incluido con el SO del Cluster. Completado este procedimiento se puede ya proceder con la instalación de las TN. En este punto se recomienda completar la instalación de al menos una de ellas a la que llamaremos en lo sucesivo la *terminal-nodo de prueba* o *TNP*. El procedimiento descrito en los parágrafos anteriores puede repetirse hasta que la TNP quede instalada y configurada como se espera queden la totalidad de las TN de la Sala Cluster.

Rocks permite además la configuración de otras características importantes como lo son la contraseña del administrador (root), la zona horaria del sistema, la

configuración del idioma del teclado, el mouse y el escritorio por defecto que tendrán los TN, e incluso tiene la opción de añadir comandos para que se ejecuten en el sistema una vez que la instalación se haya terminado. Esto último puede utilizarse con el fin de configurar software adicional. Estos lineamientos se encuentran en la documentación de Rocks (http://www.rocksclusters.org) y hacen parte del sistema de *Kickstart* y los archivos configurables en el CPD.

En la figura 2 se representa esquemáticamente en un diagrama de flujo los pasos básicos para la instalación completa de la Sala Cluster usando el procedimiento descrito en este trabajo.

3.4. Preparación de la instalación del SO utilitario en las TN

La instalación del SO del cluster usa mecanismos que la hacen eficiente y escalable y que exigen el mínimo de intervención por parte del administrador [9]. En el caso del SO utilitario estas condiciones son más difíciles de satisfacer. La instalación sistemática de Windows XP en las TN de la sala puede realizarse recurriendo a mecanismos como la instalación desde imágenes o la instalación por red. Por la naturaleza del procedimiento explorado en este trabajo recomendamos el uso de la instalación desde imágenes.

El primer paso para la instalación del SO utilitario es el de preparar una instalación completa del mismo en la TNP usando un método tradicional. Debe ponerse particular atención en los detalles de esta instalación porque todas las terminales de la sala tendrán exactamente la misma configuración, servicios, restricciones y software de terceros instalados en la TNP.

Una vez completada la instalación del SO utilitario en la TNP tenga en cuenta, que el sistema de arranque es sobre escrito por el nuevo SO utilitario. En GNU/Linux es posible utilizar una herramienta de "rescate" (*Rescue Disk*) para reconfigurar el gestor de arranque incluyendo ahora el nuevo SO. En Rocks es necesario sin embargo reinstalar completamente el SO del cluster en la TNP de modo que se actualice toda la información relacionada con la partición del nuevo sistema operativo en el *Master Boot Record* (*MBR*). Esto no representa una perdida excesiva de tiempo dada la eficiencia de los sistemas de instalación de Rocks [9]. Para garantizar que en esta etapa la partición sobre la que reposa el recién instalado SO utilitario no sea formateada se debe borrar de la base de datos de particiones que reposa en el SFE la información de particionado de ese TN, utilizando para ello el comando *rocks-partition*. Una vez se ha corregido el gestor de arranque, se puede reiniciar la TNP en GNU/Linux y proceder con la preparación de la imagen del recién instalado SO utilitario.

Para ello se puede utilizar una de las herramientas disponibles para la creación y manipulación de imágenes de particiones. Entre estas resaltan por su versatilidad las herramientas *Norton Ghost, Partition Saver, Partimage o NTFS Clone*. En la elección de la herramienta de preparación de imágenes debe tenerse presente la compatibilidad con sistemas de archivos NTFS y otras propiedades que garanticen el clonado e instalación de este tipo de sistemas de archivos. En lo sucesivo asumiremos el uso de *NTFS clone* que satisface estas condiciones y además es liviana y sencilla de utilizar.

La imagen creada con el procedimiento anterior deberá ser posteriormente copiada y almacenada en el SFE o copiada y almacenada en un disco secundario conectado temporalmente a la TPN (se recomienda la primera opción.) Desde allí (desde el SFE o desde el disco secundario en la TPN) la imagen será copiada a las demás TN en las siguientes fases de instalación.

3.5. Instalación del SO del cluster en las TN

La instalación de las TN de la sala cluster seguirá en lo sucesivo un procedimiento idéntico y bien documentado en los manuales de la distribución. En este punto disponer de mecanismos de arranque por red en las TN (PXE por ejemplo) puede hacer muy eficiente y rápido el proceso. Debe tenerse presente en este punto las limitaciones ofrecidas por la red de datos que conecta los equipos en la sala. Si bien una distribución como Rocks garantiza un flujo sostenido y regular de datos en la instalación de múltiples maquinas sobre la misma red, cuando el número de nodos es muy grande el proceso de instalación puede tomar un tiempo muy grande.

3.6. Instalación del SO utilitario en las TN

En el esquema concebido en este trabajo se debe instalar en este punto el SO utilitario cuya imagen preparamos en la etapa descrita en 3.4. La imagen primero debe copiarse del SFE (en el esquema recomendado en 3.4) a cada una de las TN una vez se ha instalada en ellas el SO del cluster. Debido al gran tamaño que puede tener la imagen, es posible utilizar, de acuerdo al espacio disponible en los discos de la TN, dos esquemas de copiado: 1) copiar de la imagen a la partición */state/partition1* del TN y después desde allí instalar la imagen en la partición destinada para el SO utilitario usando las herramientas de NTFS clone. 2) usar directamente una conexión ssh para descomprimir la imagen usando NTFS clone desde el SFE en el TN tal y como es descrito en la documentación de la herramienta de clonado.

Estos dos pasos en apariencia sencillos pueden ofrecer en algunos casos ciertas dificultades que comprometen la eficiencia del proceso. De una parte la transferencia de una imagen de algunos *Giga bytes* a través de una red convencional puede tomar (dependiendo de las características del hardware de interconexión) varias decenas de minutos. En los esquemas propuestos arriba el proceso puede siempre automatizarse. Para el caso de reparación del SO de una TN después de que la sala cluster esta completamente instalada, el mecanismo descrito aquí puede programarse para que se realice horas de menor actividad (noche o un fin de semana). En cualquier caso el tiempo que toma este tipo de procedimientos puede ser comparable al que toma la instalación convencional de una sala con la ventaja de que si se configura en forma apropiada puede ocurrir de manera automática sin la intervención de un operario.

Otra forma efectiva de instalar el SO utilitario en las TN puede ser utilizando un procedimiento "invasivo". La imagen que se creo en el TNP es almacenada en un disco secundario temporalmente conectado a ese nodo. El disco entonces se conecta a cada TN y la imagen se implanta usando las herramientas de clonado instaladas previamente en el SO del cluster, pero ahora desde un dispositivo local con un tiempo de transferencia muchas veces menor al requerido en la instalación por red. Este procedimiento es sin embargo poco recomendado por requerir siempre la intervención directa del administrador. Sólo se debería utilizar cuando se tengan restricciones importantes con la red o con la capacidad de almacenamiento de los discos en las TN o en el SFE.

Una vez instalada la imagen del SO utilitario se deberá cambiar el tipo de sistema de archivos de la partición */utilos* en la tabla de particiones (ver tabla 2). Originalmente se había configurado esta partición para almacenar un sistema de archivos VFAT. Ahora debe configurarse como del tipo NTFS. Esto se logra usando el comando *fdisk*. Hecho esto y para garantizar la coherencia de la configuración del MBR se sugiere reinstalar el SO del cluster en la TN, cuidándose previamente de borrar la información sobre el particionado en la base de datos del SFE como se hizo durante la instalación del TNP (sección 3.3). Si bien podría existir un mecanismo más eficiente que la reinstalación, este proceso es tan rápido que en las experiencias reportadas aquí no se exploraron mecanismos alternativos.

4. Recomendaciones especiales

Se presentan a continuación algunas recomendaciones especiales relacionadas con la configuración de la Sala Cluster. Se sugiere prestar especial cuidado a estas recomendaciones para garantizar de un lado que la plataforma preste adecuadamente el servicio dual para el que fue instalada y de otro para facilitar la administración de la misma (ver documentación de la sala cluster http://urania.udea.edu.co/facom/cluster-room).

Si bien los tiempos de uso de la sala como cluster deben ser claramente definidos en las políticas administrativas de la misma, es recomendable garantizar que aún en franjas de tiempo donde normalmente se usaría la sala con propósitos utilitarios (atención de usuarios que usan normalmente Windows) se garantice el mayor número posible de terminales iniciadas con GNU/Linux. Para ello es recomendable escoger el SO del cluster como sistema operativo por defecto en el gestor de arranque.

Si por algún motivo se reinstala el SO utilitario en algunos o todas las terminales de la sala debe recordarse con atención que después de dicha instalación antes de proceder con la reinstalación del SO del cluster (como se explico en la sección 3.3 y en la Figura 2) se debe borrar la configuración del particionado de la(s) terminal(es) en el SFE.

Una de las grandes bondades del sistema de instalación de Rocks es que la reinstalación del SO del cluster no destruye ni modifica la partición o el sistema de archivos del SO utilitario de modo que se pueden hacer grandes cambios al primero (actualización a versiones más recientes, reconfiguración de paquetes, etc.) sin afectar el segundo.

La configuración del sistema de arranque de Rocks (Grub) prevé que cuando el SO del cluster es interrumpido repentinamente (suspensión de la electricidad, apagado desde la fuente, etc.) el sistema arranca por defecto un mecanismo de reinstalación del sistema operativo. Este mecanismo ha sido diseñado para reparar de forma automática nodos en un gran cluster con tan solo apagarlos directamente de la fuente [4]. En una Sala Cluster sin embargo, donde cada terminal cuenta con periféricos detectar y corregir problemas en el SO del cluster es más sencillo, de modo para evitar molestas suspensiones del servicio de una terminal por un sencillo accidente (causado por un usuario desprevenido o por una fluctuación en el suministro de electricidad) se recomienda deshabilitar este mecanismo de reinstalación editando, como se explica en la documentación de la iniciativa, el archivo de configuración respectivo del Grub.

Se sugiere definir un conjunto de nodos dedicados para garantizar en lo posible la continuidad de algunos trabajos de cómputo sobre la plataforma más allá de los tiempos definidos para la operación como cluster de la sala.

5. Resultados y conclusiones

Los procedimientos descritos en este trabajo han sido exitosamente utilizados para el montaje, instalación, configuración y uso de salas cluster en el Instituto de Física de la Universidad de Antioquia y en la Escuela de Ingeniería de la Universidad Pontificia Bolivariana en Medellín (Colombia)

En el primero de los caso, la Sala Cluster viene siendo utilizada activamente desde junio de 2005 para la prestación de servicios para estudiantes y profesores del Instituto de Física y para la realización de actividades de capacitación permanente en las áreas de computación de alto rendimiento (*HPC*) y computación distribuida. Algunos trabajos de investigación estudiantiles han sido realizados usando la herramienta en horas de la noche y en fines de semana.

En la Universidad Pontificia Bolivariana se completó en septiembre de 2006 la instalación de una Sala Cluster que presta sus servicios a la Escuela de Ingeniería. La sala viene siendo regularmente utilizada en cursos de las carreras que se ofrecen en la Escuela y en los que se usan aplicaciones propietarias sobre el sistema operativo Windows XP (Matlab, LabView, HYSYS entre otras). La Sala Cluster además ha comenzado a prestar sus servicios como plataforma de computo distribuido a dos proyectos de investigación en las áreas de simulación en computación cuántica y microelectrónica [3] y para realizar pruebas de una infraestructura de Grid inter universitaria en la ciudad de Medellín [1]. En este caso dadas las políticas de uso de las salas vigentes en la Escuela de Ingeniería de la Universidad Pontificia Bolivariana la sala como cluster se usa solamente durante las horas de la noche y los fines de semana.

Se enumeran a continuación algunas de las preguntas y cuestiones más relevantes que enfrentan los administradores de salas de computo pública cuando se plantea la posibilidad de convertirlas en salas clusters. La solución a estas y otras preguntas se presenta de forma más detallada en la documentación de la iniciativa.

¿*Cuánto tiempo toma instalar una sala-cluster y como se compara con la instalación convencional de una sala solo con el sistema operativo Windows*? El procedimiento de instalación de la Sala Cluster es naturalmente más intrincado (al principio) que la instalación de una sala con un solo sistema operativo. Sin embargo dada la automatización de muchos procesos y la eficiencia de los mecanismos de instalación del SO del cluster, el tiempo que toma la instalación del sistema no es mucho mayor que una instalación normal, y mejor aún las tareas de reinstalación o actualización son mucho más eficientes. Los tiempos aproximados (y basados en la experiencia de los autores) para cada una de las fases descritas en este trabajo se presentan en la tabla 3. Como puede verse de los valores consignados en la tabla dependiendo de distintos factores la instalación puede tomar entre unas 5 horas continuas (1 jornada) a unas 15 horas o 4 jornadas aproximadamente.

¿Se modifica la experiencia del uso de la sala para los usuarios regulares de la misma? Esta, que parece una pregunta sencilla, es una preocupación común entre los administradores de salas públicas que buscan mantener un alto nivel de calidad de servicio en las instalaciones. La experiencia de los autores en este sentido ha sido diversa. Todo depende de las políticas definidas para el uso de la sala y para la distribución del tiempo. En la Universidad de Antioquia donde se tienen políticas de prioridad para la investigación en las salas de computo, se da prioridad y preferencia al trabajo con el sistema operativo GNU/Linux. Los usuarios tienen conciencia de ello y normalmente se acostumbran a usar GNU/Linux como terminales utilitarias reduciendo el impacto sobre los trabajos de investigación que corren en la Sala Cluster. En la Universidad Pontificia Bolivariana donde las salas están orientadas a la capacitación sobre Windows, durante los tiempos de servicio público de la sala las terminales nodos están encendidas permanentemente en Windows y no hay ninguna diferencia con una sala convencional desde la perspectiva del usuario. Al iniciar la jornada en la mañana o al terminar el fin de semana se deben reiniciar las máquinas nuevamente en Windows pero este procedimiento es equivalente al de encender las máquinas en una sala convencional.

¿*Cómo se modifica la vida útil de los equipos de la sala al estar encendidos en horarios en los que no se presta servicio público*? Naturalmente la vida de útil de los equipos y sus partes se reduce en un esquema de uso prácticamente continuado. Sin embargo este hecho se ve compensado por el incremento en la eficiencia con la que son utilizados los equipos que ahora prestan un servicio a un número mayor de usuarios pero también para un número mayor de aplicaciones. En la Universidad de Antioquia, donde la Sala Cluster ha operado por un tiempo más prolongado, el número de incidentes de mantenimiento por año (falla de partes) ha sido mayor comparativamente con el de otras salas en el mismo período. Sin embargo la sala se ha convertido en ese mismo tiempo en un ejemplo de utilización eficiente de los recursos de cómputo de una dependencia. Además nuevos cursos específicos en las áreas de computación distribuida y computación de alto rendimiento se han ofrecido en la sala como único espacio disponible para hacerlo, lo que ha atraído la inversión en la misma desde la administración del instituto y de otras dependencias.

¿Necesitan los administradores y auxiliares una formación y capacitación en computación distribuida y de alto rendimiento para instalar y operar la sala?

En general no. Las salas clusters instaladas en la Universidad de Antioquia y en la Universidad Pontificia Bolivariana han sido administradas durante casi dos años por auxiliares temporales que con solo unas cuantas horas de instrucción aprenden a manejar los procedimientos operativos básicos de la sala. Para la instalación se requiere conocimientos específicos en el uso e instalación de GNU/Linux, habilidades cada vez más comunes entre los administradores. Adicionalmente nuestro equipo viene desarrollando algunos aplicativos y scripts para facilitar la administración de la sala como cluster por administradores sin conocimientos o experiencia en el área.

Inicialmente se espera que los procedimientos descritos en este trabajo puedan utilizarse para el montaje de clusters parcialmente dedicados, con una baja inversión económica en otras universidades en Colombia. Como una muestra del impacto que esta iniciativa tendría en la inserción de investigadores y estudiantes en el uso de computación distribuida y computación de alto desempeño se están realizando esfuerzos para su inclusión en la iniciativa de constitución de un Grid Nacional que conectará Clusters Universitarios a través de la Red Nacional Académica de Tecnología Avanzada. Algunos participantes de la iniciativa, que es soportada por la Agenda de Conectividad, RENATA, el Ministerio de Educación y Colciencias, han empezado a considerar el uso de la metodología para el montaje de sus propios Clusters.

6. Referencias

[1] A. Ospina, J. Zuluaga, “Pruebas de conectividad y procesamiento distribuido con Clusters Beowulf usando la Red Regional de Antioquia de Tecnología Avanzada (RUANA),” Proyecto de investigación, 2007, CIDI, UPB, Código 092A-06/07-20.

[2] C. Boehme, T. Ehlers, J. Engelhardt, A. Felix, O. Haan, T. Kalman, B. Neumair, U. Schwardmann, D. Sommerfeld, "Instant-Grid: Fully Automated Middleware-Deployment Using a Live-CD," icns, p. 70,2006.

[3] C. Peñuela, A. Ospina, J. Montoya, “Simulación de un Computador Cuántico, utilizando procesamiento en paralelo,” Proyecto de investigación, 2006-presente, CIDI, UPB, Código 910-05/06-20.

[4] Greg Bruno, Comunicación personal con J. Zuluaga, 2007.

[5] Greg Bruno, Mason J. Katz, Federico D. Sacerdoti, and Philip M. Papadopoulos, “Rolls: Modifying a Standard System Installer to Support User-Customizable Cluster Frontend Appliances,” IEEE International Conference on Cluster Computing, San Diego, September 2004.

[6] H. Masuda, A. Saitoh, M. Nakanishi and S. Yasutome, “Diskless Linux system with unionfs for an educational computer center,” Proc. 33rd ACM Symp. SIGUCCS conference on User services, pp. 22-26, 2005.

[7] J. Zuluaga, A. Ospina “Consideraciones metodológicas para Instalación y Configuracion de una Sala Cluster,” Revista OMEGA, No. 19, ISSN 1692-0872, 2007.

[8] J. Zuluaga, D. Mejía. “Instalación y Uso de una Sala Cluster usando NPACI Rocks (partes 1, 2 y 3).” Taller presentado durante el Encuentro de Investigación sobre Tecnologías de la Información Aplicadas a la Solución de Problemas, EITI2005: “El uso del paralelismo en las soluciones informáticas,” ISBN 958655895-0, (2005).

[9] Philip M. Papadopoulos, Caroline A. Papadopoulos, Mason J. Katz, William J. Link, and Greg Bruno, “Configuring Large High-Performance Clusters at Lightspeed: A Case Study,” Clusters and Computational Grids for Scientific Computing, December 2002.

[10] Philip M. Papadopoulos., Mason J. Katz, and Greg Bruno, “NPACI Rocks: Tools and Techniques for Easily Deploying Manageable Linux Clusters,” Concurrency and Computation: Practice and Experience Special Issue: Cluster 2001.

[11] Russell Carter, John Laroco, "Commodity Clusters: Performance Comparison Between PC's and Workstations", Symposium on High Performance Distributed Computing, p. 292, 1996.

[12] Sarah M. Diesburg, Paul A. Gray, David Joiner, "High Performance Computing Environments Without the Fuss: The Bootable Cluster CD", International Parallel and Distributed Processing Symposium- Workshop 13, p. 252a, 2005.

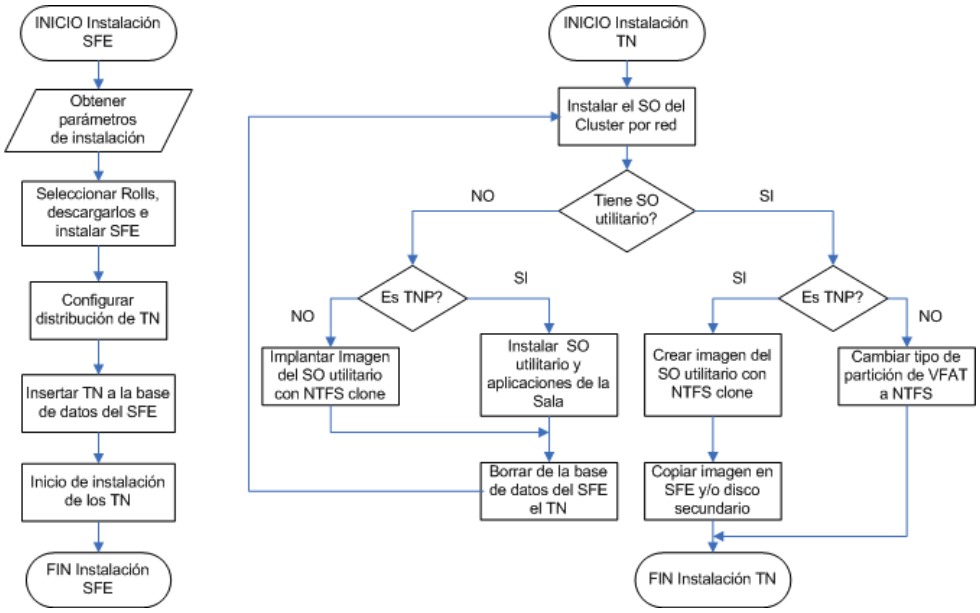

Figura 2. Diagrama de flujo del proceso de instalación de la Sala Cluster. Particular atención se debe prestar en la toma de decisiones (condicionales) en el proceso

Tabla 3. Estimado de los tiempos invertidos en la instalación de una sala-cluster y una sala tradicional.

Fase	*Tiempo estimado*	*Instalación tradicional*	*Tiempo estimado*
Instalación del SFE			
Instalación y configuración del SFE	50 m - 1 h 10 m	Instalación servidor de dominio MS	1 h - 4 h
Instalación del TNP			
Instalación del SO del cluster	20 m	No aplica	
Instalación del SO utilitario	30 m – 4 h	Instalación de SO en una terminal	30 m – 4h
Creación de la imagen del SO utilitario	10 m – 1 h 20 m	Creación SO utilitario	10 m – 1 h 20 m
Re instalación SO del cluster	20 m	No aplica	
Instalación del TN			
Instalación SO en TNs	30 m – 50 m[1]	No aplica	
Transferencia de la imagen a los TN	50 m[2] – 4 h 10 m[3]	Transferencia de la imagen	50 m – 4 h 10 m
Reinstalación SO del cluster en TNs	30 m – 50 m	No aplica	
Configuración del SO utilitario	1 h 40 m	Configuración de las terminales	1 h 40 m
TOTAL	**5h40 m – 14h30m**		**4h10 m – 15h10m**

[1] *Se comparte la red durante la instalación,*
[2] *Suponiendo una transferencia invasiva a un total de 10 TNs (5 m/TN),*
[3] *Suponiendo una transferencia por red compartida por 4 nodos por vez (1h 40m / 4 TN)*