



Adaptación de Clusters de Linux para Servicios de Redes

Linux Clusters Adaptation for Network Services

Gilberto Díaz

Departamento de Computación, Escuela de Sistemas, Facultad de Ingeniería
Universidad de Los Andes, Mérida 5101, Venezuela
gilberto@ula.ve

Chaves J., Mendoza V.

Centro de Cálculo Científico (CeCalCULA)
Universidad de Los Andes, Mérida 5101, Venezuela
juanluis@ula.ve, victorm@ula.ve

Hoeger H.

Centro de Simulación y Modelos, Facultad de Ingeniería
Universidad de Los Andes, Mérida 5101, Venezuela
hhoeger@ula.ve

Núñez L.

Centro de Astrofísica Teórica, Departamento de Física, Facultad de Ciencias
Universidad de Los Andes, Mérida 5101, Venezuela
nunez@ula.ve

Resumen

Los clusters o agrupaciones de PCs constituyen una técnica para construir una máquina paralela virtual a partir de computadores personales y software libre, con prestaciones comparables a los de los supercomputadores tradicionales. La redundancia natural de los clusters y su gran escalabilidad los convierten en una plataforma muy adecuada para otras áreas como los servicios de redes (correo electrónico, alojamiento de sitios web, etc.). Sin embargo, no existen herramientas de balanceo de carga en software libre apropiadas para clusters de servicios de redes como sucede en clusters para computación de alto rendimiento. De acuerdo al tipo de servicio de red que se implante en un cluster de PCs, se debe utilizar una técnica de balanceo de carga distinta. Este artículo describe las técnicas de balanceo de carga utilizadas para adaptar los cluster Linux a los servicios de redes de la Universidad de Los Andes: el correo electrónico y alojamiento de sitios web.

Palabras Claves: Clusters, GNU/Linux, Computación de Alto Rendimiento, Correo Electrónico, Alojamiento Web

Abstract

PC Clusters are a technique to build a virtual parallel machine using common personal computers and free software (open source) with similar performance than traditional supercomputers. Natural redundancy and big scalability make them a very suitable platform for network services (e-mail, web hosting, etc). However, there are not proper open source load balancing tools for network services clusters like in High Performance Computing clusters. According to the network service implanted in a cluster we must use a specific load balancing technique. This article describe the load balancing strategies used to adapt Linux Clusters for Los Andes University network services: electronic mail and web hosting.

Key Words: Clusters, GNU/Linux, High Performance Computing, Electronic Mail, Web Hosting





1. Introducción

En los últimos años la incorporación de avances de arquitectura en los microprocesadores ha sido significativa. Sin embargo, no podemos depender continuamente de procesadores más rápidos para obtener más eficiencia. Hay límites físicos, como la velocidad de la luz, que eventualmente van a desacelerar la reducción, que se ha visto año tras año, en el tiempo que dura un ciclo (tiempo para ejecutar la operación más básica) de CPU. Dadas las dificultades en mejorar la eficiencia de un procesador, la convergencia en eficiencia entre microprocesadores y los supercomputadores tradicionales, y el relativo bajo costo de los microprocesadores¹, ha permitido el desarrollo de computadores paralelos viables comercialmente con decenas, cientos y hasta miles de microprocesadores. Un *computador paralelo* es un conjunto de procesadores capaces de cooperar en la solución de un problema. Esta definición incluye supercomputadores con cientos de procesadores, máquinas con múltiples procesadores, redes de estaciones de trabajo (NOWs²) y redes de PCs (*Clusters* de PCs).

En los 90 se dieron ciertos eventos muy favorables en el área. Los PCs comienzan a exhibir la capacidad de las estaciones de trabajo, sus precios se hacen muy asequibles y los costos de los equipos de redes disminuyen significativamente. Por otro lado, el surgimiento de *LINUX*, un sistema operativo completamente abierto originalmente desarrollado por el Finlandés *Linus Torvalds* y luego mediante la colaboración de un sin número de voluntarios alrededor del mundo, compatible con *UNIX* y capaz de correr sobre PCs, permite finalmente satisfacer las demandas de computación a una fracción del costo asociado a los supercomputadores.

Los clusters o agrupaciones de PCs nacen de la necesidad de tener una plataforma alterna a los grandes y costosos sistemas de supercómputo para el desarrollo de aplicaciones paralelas y la investigación científica. Donald Becker y Thomas Stirling en 1994 crearon una técnica para construir una máquina paralela virtual a partir de una red de computadores personales para un proyecto de la NASA. A este sistema lo denominaron *Beowulf* y se convirtió en el modelo de los *clusters* de PCs para la computación de alto rendimiento. Así mismo, la calidad de los servicios

¹ Tienen una demanda sustancialmente mayor que la de otros procesadores que permite dividir los costos de diseño, producción y comercialización entre más unidades.

² Network of Workstations.

de redes en la actualidad, además de las prestaciones de los enlaces, depende fuertemente del rendimiento de los servidores. Las características intrínsecas de los arreglos de PCs los hacen muy apropiados como plataforma de alto rendimiento, servidores web, servidores de correo, almacenamiento, etc.

En el área de Computación de Alto Rendimiento existen distintas herramientas que sirven para que este conjunto de máquinas se comporten como un único computador paralelo y que se encargan de balancear la carga entre ellas. Sin embargo, no hay herramientas estándares y apropiadas para balancear carga en servicios de redes.

El Centro de Cálculo Científico de la Universidad de Los Andes (CeCALCULA) utiliza esta técnica desde 1997 para brindar apoyo a los distintos investigadores de la ULA, de instituciones académicas de Venezuela y otros países. La experiencia lograda en este centro de cómputo sirvió como fundamento para adaptar la técnica de agrupar computadores personales a los servicios de redes tales como: correo electrónico, alojamiento de sitios web, etc. De esta manera, el correo electrónico de la Universidad de Los Andes se fundamenta desde el 2003 en un cluster Linux brindando una plataforma que gestiona actualmente más de 5 millones de mensajes trimestrales. Desde entonces este servicio ha podido enfrentar la creciente demanda de los usuarios y ha proporcionado un mecanismo que escala adecuadamente ante situaciones que exigen la atención de nuevos requisitos.

2. ¿Por Qué Clusters de PCs?

Las computación desde sus inicios se convirtió en una alternativa indispensable para los métodos tradicionales de la investigación científica[2]: análisis teórico y experimentos de laboratorio. El enfoque computacional frecuentemente ofrece soluciones a problemas que por su complejidad son prácticamente imposibles de manejar desde el punto de vista teórico. Así mismo, el computador puede proporcionar información que no está disponible en el laboratorio y también puede proveer nuevas formas de análisis a los problemas estudiados. El grupo de profesionales que mantiene el sitio *top500*³ ofrece información sobre las quinientas (500) máquinas más rápidas del planeta utilizadas en investigaciones de diferente índole. Ellos utilizan *linkpack*⁴ un benchmark que consiste en un sistema denso de ecuaciones lineales. Las estadísticas de Junio del 2007 del *top500* muestran las diferentes

³ <http://www.top500.org>

⁴ <http://www.netlib.org>





arquitecturas más utilizadas actualmente para la Computación de Alto Rendimiento (gráfico de la figura 1)

campo según las últimas estadísticas presentadas en el top500. El gráfico de la figura 2: Arquitecturas más utilizadas en el top500 ilustra esta afirmación.

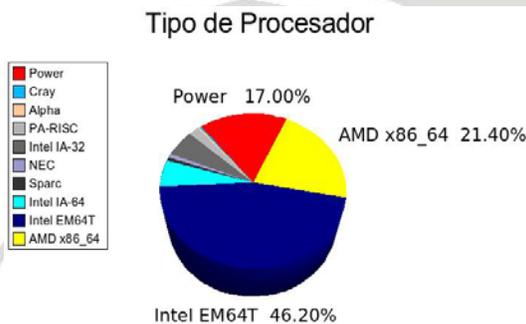


figura 1: Procesadores del top500.org

Arquitecturas como POWER y Cray ofrecen plataformas ideales para la computación de alto rendimiento las cuales lideraron en la década de los 90. Sin embargo, como se observa en el gráfico anterior actualmente son poco utilizadas y las tecnologías Intel EM64T y AMD x86_64 son preferidas para la computación de alto rendimiento. Los actuales procesadores Intel de 3GHz tienen capacidad de hasta 12 GFLOPs y en los QuadCore podemos tener hasta 48 GFLOPs. Capacidades similares se pueden obtener con procesadores AMD.

En los 90 los PCs comienzan a exhibir la capacidad de las estaciones de trabajo y sus precios se hacen muy asequibles. Así mismo, en el área de redes de computadores, el rendimiento de los equipos de comunicación y sus bajos costos permiten implantar redes en muchas instituciones, particularmente en universidades y otros institutos académicos. Por otro lado, el surgimiento de LINUX, un sistema operativo gratuito originalmente desarrollado por el Finlandés Linus Torvalds y luego mediante la colaboración de un sin número de voluntarios alrededor del mundo, compatible con UNIX y capaz de correr sobre PCs, permite finalmente satisfacer las demandas de computación a una fracción del costo asociado a los supercomputadores. Todo esto, ofreció una serie de condiciones favorables para que en 1994 Donald Becker, Thomas Stirling y su equipo en la NASA, utilizaran una técnica para agrupar un conjunto de PCs, logrando obtener una eficiencia comparable a los supercomputadores. A este sistema lo denominaron Beowulf y se convirtió en el modelo de los clusters de PCs para la computación de alto rendimiento.

Actualmente estas condiciones han influenciado significativamente la forma de hacer computación de alto rendimiento y los clusters de PCs dominan el

Arquitecturas

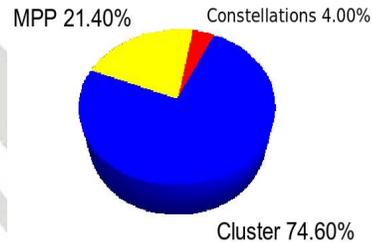


figura 2: Arquitecturas más utilizadas en el top500.org

Además de presentar un rendimiento comparable al de máquinas más sofisticadas a una fracción de su costo, los clusters de PCs poseen otras ventajas que los hacen muy apropiados para la Computación de Alto Rendimiento:

- **Ensamblaje:** no se requiere tener un doctorado en computación y años de experiencia para ser capaz de construir un cluster. Hoy en día estudiantes de bachillerato son capaces de ensamblar PCs. Las partes se pueden comprar por separado: tarjeta madre, procesador, tarjeta de video, tarjeta de sonido, disco duro, lectora/escritora de CDs, monitor, teclado, fuente de poder, etc., de acuerdo a los gustos y necesidades, y al acoplarlas tienen un PC hecho a la medida. Con ciertos conocimientos adicionales de redes pueden armar un cluster.
- **Mantenimiento:** dado que los elementos que forman un cluster se encuentran fácilmente en el mercado (son componentes de producción masiva y por lo tanto de bajo costo), al fallar alguno de ellos se puede reemplazar sin mayores inconvenientes.
- **Alta disponibilidad:** Si algún componente falla, solamente el nodo asociado permanece fuera de servicio. El resto de las PCs pueden atender solicitudes provenientes de los clientes. Existen mecanismos para evitar el envío de solicitudes a los nodos que tienen fallas y así evitar que un cliente piense que el servicio ya no está disponible.
- **Tolerancia a fallas:** Los clusters están formados por PCs individuales interconectados por una red y en la gran mayoría de los casos no es necesario poner fuera de servicio todo el cluster para reemplazar un componente, sino solo el nodo (el PC o máquina) al que está asociado el





componente. Por lo general los servidores especializados constan de CPUs interconectados por redes especiales dentro de una misma caja. Reemplazar un componente implica apagar totalmente la máquina. Además hay que esperar que llegue el experto de la compañía con la pieza muy particular y costosa para reparar el supercomputador.

- **Hospedaje:** debido al alto costo de los servidores especializados, estos deben ser albergados en centros especiales. Dentro de estos centros, ellos se encuentran en salas muy particulares con sistemas de aire acondicionado, filtrado de aire, ductos de enfriamiento, cableados y sistemas de protección eléctrica especiales, etc. Además de que deben contar con administradores, consultores, personal de mantenimiento, etc., con una preparación importante en el manejo de estas máquinas. Los clusters requieren de alojamientos mucho más modestos con el requerimiento principal de poseer un sistema eléctrico adecuado. En muchos casos ni siquiera hace falta poseer aire acondicionado.
- **Modernización y expansión:** por la esencia misma de lo que son los servidores especializados, cuando un centro recibe una de estas máquinas dentro de poco aparecen nuevos modelos. Actualizar los supercomputadores se traduce en comprar los nuevos modelos. Por lo particular que son los supercomputadores, expandir sus capacidades de memoria, almacenamiento en disco, número de CPUs, etc., se traduce en inversiones sustanciales. Como los clusters están compuestos por elementos disponibles de múltiples fabricantes y debido a la compatibilidad que estos tratan de mantener con las diferentes generaciones de una misma familia de componentes, se hace sencillo modernizarlos. Actualizar el cluster con CPUs más potentes puede ser tan sencillo como sacar un CPU de la tarjeta madre e instalar otro, o quizás, reemplazar la tarjeta madre y el CPU conservando el resto de los componentes: memoria, tarjetas de video, etc. Expandir la capacidad de memoria y de almacenamiento en disco no requiere de inversiones sustanciales dado el bajo costo de estos componentes. Añadir CPUs implica agregar PCs de fácil adquisición.
- **Escalamiento:** Si la demanda de un servicio crece, no es necesario desechar la plataforma existente para incorporar máquinas más poderosas y así aumentar las prestaciones del servicio. Agregar más nodos al cluster es muy fácil, estos pueden tener mejores prestaciones que los nodos actuales

o simplemente tener las mismas características. En cualquier caso, se incrementa la capacidad de atención de solicitudes del servicio.

Otro factor importante que influye de forma significativa es el sistema operativo. Una vez más, según el *top500* el sistema operativo más utilizado actualmente en las quinientas máquinas más rápidas del mundo el GNU/Linux. (gráfico de la figura 3: Sistemas Operativos utilizados en el top500)

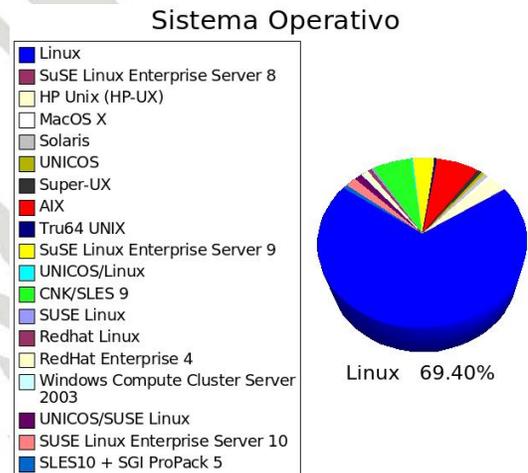


figura 3: Sistemas Operativos utilizados en el top500.org

Por todo esto, podemos decir que la mejor opción para la computación paralela y el cálculo intensivo son los clusters Linux. En el **Centro de Cálculo Científico de la Universidad de Los Andes (CeCaLCULA)** se ha utilizado la tecnología de clusters Linux desde 1997 para dar apoyo a los científicos en diferentes áreas de la investigación. En los inicios se contaba con dos pequeños clusters:

- **Atma:** Cluster de cinco nodos dedicados a la ejecución de modelos de Monte Carlo
- **Chama:** Cluster de ocho nodos de propósito general para ejecución de aplicaciones de física y química computacional. En el 2000 este creció a 20 nodos incrementando su capacidad de procesamiento y almacenamiento.
 - Paralelismo
 - Química computacional
 - Física computacional
 - Bioinformática
 - Arquitectura computacional

El adiestramiento en clusters Linux siempre ha sido un objetivo fundamental de CeCaLCULA. Con éste se persigue la difusión del uso de clusters, tanto para usuarios finales como para administradores, con la finalidad de que la comunidad científica nacional y





regional (países andinos y del Caribe) cuentan con esta herramienta para resolver problemas de mayor dimensión. En 1998 se realizó un taller con estudiantes de distintas universidades del país (US, LUZ, ULA y UNET) donde se generó un documento que describía en detalle la instalación, configuración y uso de un cluster⁵. En ese mismo año se repitió la experiencia en el ICIMAF (Habana – Cuba) con científicos de distintas instituciones de ese país. En el 2003 se realizó la Primer Taller Latinoamericano de Computación de Alto rendimiento en Clusters Linux (ULA – Mérida – Venezuela), con el auspicio del International Centre of Theoretical Physics (ICTP – Trieste – Italia). En el 2005 este taller se llevó a la Universidad Industrial de Santander (UIS – Bucaramanga – Colombia) y al CENATAV (Habana – Cuba). En el 2006 se dictó un breve taller en el Centro Universitario de Ixtlahuaca (CUI – Ixtlahuaca – México). En el 2007 se le dió adiestramiento en clusters Linux a investigadores del Instituto de Geofísica del Perú (IGP – Lima – Perú). Así mismo, una serie de talleres locales para científicos de distintas áreas han sido impartidos en los últimos 10 años.

3. Demanda Intensiva de Servicios

En la década de los 90 todo tipo de organizaciones iniciaron la migración de muchos de sus servicios a sus versiones electrónicas, más aún, actualmente muchos de ellos sólo son alcanzables de esta manera. Los servidores Web son el mecanismo más común para ofrecer estos servicios electrónicos. El acceso a éstos, es más fácil y expedito que los servicios tradicionales de difusión de información. Esta condición, propicia que un número mucho mayor de usuarios pueda hacer uso de ellos. Por otro lado, el rápido crecimiento de Internet y su expansión proporciona acceso desde cualquier parte del mundo a los servicios en línea.

Todo esto hace posible que el número de solicitudes por unidad de tiempo se incrementa de forma sustancial. Por esto, un servicio debe contar con una plataforma que pueda satisfacer cantidades ingentes de accesos que sea capaz de satisfacer la creciente demanda y escalar a la par de las nuevas exigencias.

Por otro lado, la cantidad de información por solicitud se ha incrementado significativamente. Un caso particular es el tamaño de los adjuntos de los mensajes de correo electrónico. El gráfico de la figura 4: Tamaño promedio de mensajes muestra los registros del tamaño promedio de los mensajes gestionados por

los servidores de correo electrónico de la Universidad de Los Andes. En este se puede observar claramente como el tamaño promedio de los correos ha crecido en los últimos años desde 30KB hasta casi 200KB.

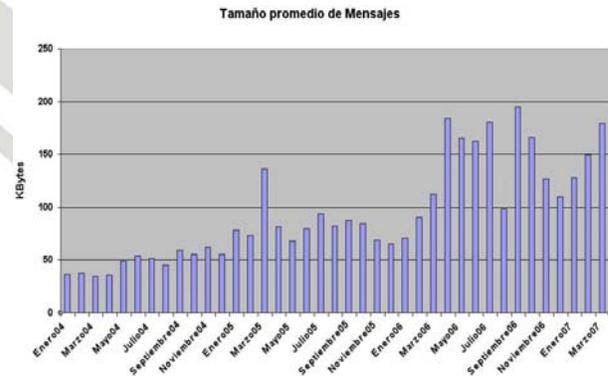


figura 4: Tamaño promedio de mensajes

Otro problema que deben enfrentar los servicios de redes hoy en día es la carga generada por solicitudes no válidas como por ejemplo flujos ingentes de paquetes que pueden degradar el servicio (ataques DoS⁶ y virus⁷) Para ejemplificar esto utilizamos una vez más la plataforma de correo de la Universidad de Los Andes. En el gráfico de la figura 5: Correo no Deseado se puede observar como el número de mensajes no deseados (spams⁸ y adjuntos infectados con virus) ha crecido sustancialmente en los últimos años.

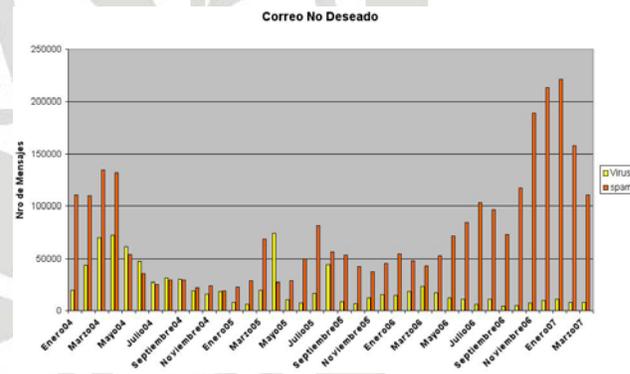


figura 5: Correo no Deseado

De igual manera que en la computación de alto rendimiento, en los servicios de redes, una sola

⁵ <http://www.cecalc.ula.ve/manuales/clusters.pdf>

⁶ DoS: Denied of Service o negación del servicio son ataques comunes cuya finalidad es sobre cargar a los servidores con un flujo intenso de solicitudes falsas.

⁷ Algunos virus o gusanos informáticos generan grandes cantidades de solicitudes para duplicarse e infectar nuevos equipos.

⁸ Spam: mensajes de correo electrónico no solicitado y que son enviados por lotes.





máquina no puede satisfacer una gran demanda del servicio. La estrategia utilizada en la computación paralela es dividir un problema grande en trozos más pequeños y asignar cada uno de ellos a distintos procesadores para reducir el tiempo total de cómputo. La Red de Datos de la Universidad de Los Andes (ReDULA) ha juntado esfuerzos con CeCalCULA para extender la experiencia de los clusters Linux a los servicios de redes. Un grupo de técnicos interdisciplinarios aplicó estas técnicas al correo electrónico de la ULA, y la estrategia utilizada fue dividir la atención de todas las solicitudes entre un grupo de servidores similares. De esta forma, la capacidad de respuesta creció y un número mayor de solicitudes pueden ser atendidas sin sufrir sobrecargas tempranas del servicio. Desde el 2003 se utiliza un cluster Linux para dar el servicio de correo electrónico en la ULA y podemos decir que la plataforma ha satisfecho las necesidades del servicio. La arquitectura que se utilizó para este servicio se muestra en la figura 6: Configuración Actual.

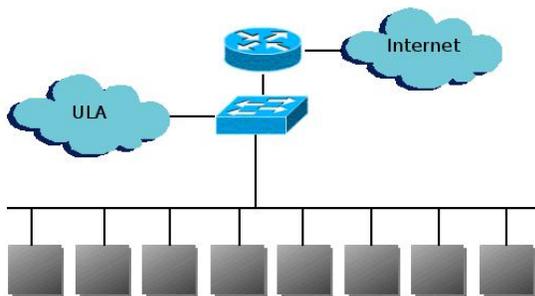


figura 6: Configuración Actual

Este cluster atiende a diez mil cuentas de correo electrónico y alberga los siguientes servicios:

- Servicios de Almacenamiento (buzones)
- Webmail
- Listas de correo electrónico
- Servicio de correo (SMTP)
- Servicio de courier (POP, IMAP)
- Servicio de directorios (LDAP)
- Mecanismos de detección de correo no deseado (antispam y antivirus)

Para atender a toda la comunidad universitaria (10 mil empleados y 45 mil estudiantes) se implantará una plataforma que divida los distintos servicios en clusters distintos para satisfacer la creciente demanda y soportar las nuevas formas de mensajes no deseados. Esta arquitectura se muestra en la figura 7: Solución Propuesta. El **Cluster A** atenderá a los servicios que el usuario utiliza directamente (Webmail, POP e IMAP), el **Cluster B** tendrá los mecanismos de detección de

mensajes no deseado. El **Cluster C** mantendrá los servicios de directorios, el **Cluster D** servirá de almacenamiento y el **Cluster E** atenderá el flujo entrante y saliente de mensajes para agregar un mejor nivel de seguridad a la plataforma.

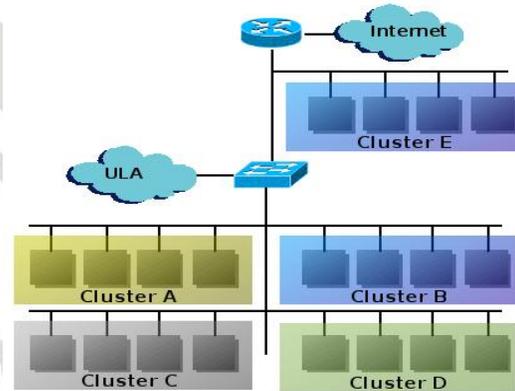


figura 7: Solución Propuesta

Gracias a la experiencia positiva desarrollada con el correo electrónico, se está implantando una plataforma similar para atender los servicios Web de la ULA. Esta tendrá un grupo de nodos para el almacenamiento, otro para los servicios de bases de datos y finalmente un grupo de nodos de interfaz con los clientes.

4. Balanceo de Carga

En los cluster para Computación de Alto Rendimiento (CAR) el problema se divide en trozos más pequeños y se asigna cada trozo a un nodo. De esta manera el problema se resuelve más rápido pues cada nodo tarda menos en resolver sólo una porción del problema. La solución general se obtiene juntando las soluciones parciales. Si es necesario resolver más de un problema a la vez entonces se debe cuidar que la carga entre los nodos se mantenga en niveles similares, de lo contrario un grupo de tareas, pertenecientes a un mismo problema, podría esperar por la finalización de una de sus integrantes por estar corriendo en un nodo más cargado. El problema de mantener una carga similar en los nodos del cluster se denomina **balanceo de carga**. En un cluster para cómputo intensivo la asignación de tareas a cada nodo se realiza a través diversos mecanismos para balancear la carga:

- Bibliotecas paralelas (MPI, PVM, etc.)
- Despachadores de tareas (PBS, SGE, etc.)
- Sistema Operativo (Módulos para el kernel de Linux)



En los servicios de redes, el balanceo de carga tiene básicamente los mismos principios pero su implantación difiere significativamente. La configuración que generalmente se utiliza en un cluster para cálculo intensivo se muestra en la figura 8: Configuración Estándar

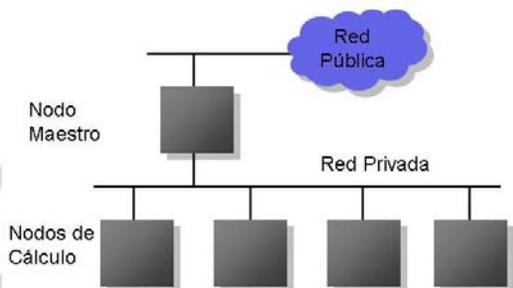


figura 8: Configuración Estándar

El nodo “maestro” se utiliza como nodo de entrada y se encarga del balanceo de carga utilizando alguno de los mecanismos antes mencionados. Desde este nodo el usuario ejecuta sus tareas y cada subtarea se envía a los nodos de cálculo manteniendo la carga balanceada.

En un servicio de redes se puede utilizar básicamente la misma arquitectura pero el balanceo de carga debe realizarse de forma diferente. En principio se reemplaza el nodo maestro por un **balanceador de carga** y los nodos de cálculo por servidores. El balanceador redirige las solicitudes provenientes de los clientes de la red pública hacia los servidores distribuyéndolas de acuerdo a la carga de cada uno. El balanceador de carga para un servicio de red puede implantarse utilizando uno de los siguientes esquemas:

- Hardware
- Software
 - Servicio de nombres (DNS)
 - Aplicaciones específicas

4.1. Balanceo de Carga por Hardware

Este primer enfoque consiste en sustituir el nodo maestro de la figura 8: Configuración Estándar por un equipo de comunicaciones especializado (Switch capa 7)⁹. Estos equipos son capaces de realizar el balanceo de carga a nivel de capa 7 y adicionalmente supervisan el estado de los servidores; si uno de estos no está

⁹ Switch capa 7: switch de red capaz de gestionar el tráfico no sólo al nivel de la capa 2 sino en las capas superiores. Se basa en los protocolos de aplicación, por ejemplo, HTTP, SMTP, etc. para realizar el encaminamiento de los paquetes.

disponible, entonces deja de enviarle solicitudes. Sin embargo, este enfoque tiene la desventaja de ser la solución más costosa.

4.2. Balanceo de Carga por Software

Este esquema utiliza distintos componentes de software para implantar las funciones realizadas por el nodo maestro de la figura 8: Configuración Estándar.

Servicio de Nombres (DNS)

Domain Name Service (DNS) es un protocolo de redes que permite la utilización de nombres (por ejemplo www.cecalc.ula.ve) para hacer referencia a sitios en Internet en lugar utilizar identificadores numéricos (direcciones IP). El balanceo de carga por DNS puede realizarse gracias a una de las funcionalidades que ofrece este protocolo la cual consiste en la asignación de un mismo nombre a distintas direcciones IP. Un servidor DNS responderá con una dirección IP distinta cada vez que un cliente solicite el identificador numérico correspondiente a un nombre en particular. La selección de la dirección IP por parte del servidor DNS es realizada utilizando un esquema **round robin**¹⁰. Entonces el cliente hará la solicitud a un servidor distinto cada vez que inicie la conexión. En la figura 9: DNS Round Robin se muestra esta estrategia.

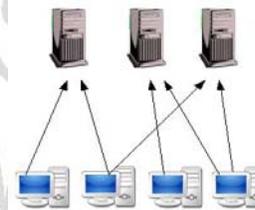


figura 9: DNS Round Robin

Este enfoque puede ser utilizado en cualquier servicio de red, sin embargo, tiene el inconveniente de no contar con mecanismos para redirigir las solicitudes ante fallas de los nodos y estas no podrán ser satisfechas como se muestra en la figura 10: Falla de un servidor.

¹⁰ Round Robin: Es un técnica que selecciona una opción entre varias alternativas de forma consecutiva. Una vez que se hace uso de la última, se comienza desde el principio.



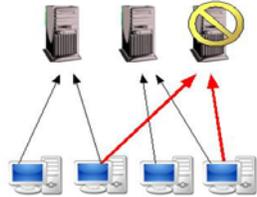


figura 10: Falla de un servidor

En servicios de correo electrónico este inconveniente tiene solución a través de los registros MX¹¹ de DNS. Estos registros se utilizan para indicar servidores de correo alternativos y los clientes pueden dirigirse a ellos en caso de no obtener una respuesta a una solicitud por parte del servidor principal. Un parámetro especial puede utilizarse para indicar el orden (prioridad) en que los distintos servidores deben ser consultados. Esto se ilustra en la figura 11: Solicitudes al servidor de más prioridad y la figura 12: Solicitudes redirigidas por falla.

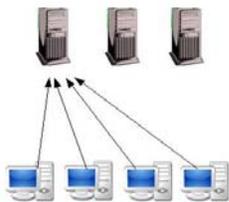


figura 11: Solicitudes al servidor de más prioridad



figura 12: Solicitudes redirigidas por falla

Este enfoque por sí sólo no funciona pues no realiza el balanceo de carga. A través de la combinación de las dos estrategias anteriores se puede obtener una solución completa. El *round robin* proporciona el balanceo de carga y los registros MX brindan la alta disponibilidad. La figura 13: DNS Round Robin y la figura 14: DNS Round Robin - MX muestran el comportamiento de esta combinación.



figura 13: DNS Round Robin

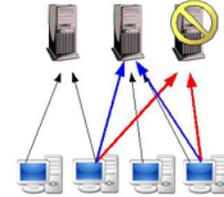


figura 14: DNS Round Robin - MX

Esta estrategia es la que se ha utilizado para implantar el cluster de correo electrónico de la Universidad de Los Andes demostrando durante los últimos 4 años un rendimiento apropiado para la creciente demanda por parte de sus usuarios.

Aplicaciones para Balanceo de Carga

En servicios de redes como servicios Web no se cuenta con mecanismos como los registros MX que proporcionen tolerancia a fallas para complementar el balanceo de carga y por eso es necesario una implantación diferente. El nodo maestro de la debe utilizar una aplicación que sea capaz de encaminar las solicitudes a un grupo de servidores balanceando la carga y no hacer asignación alguna a aquellas máquinas que no estén disponibles. Como ejemplo de este tipo de aplicaciones podemos hacer mención de **Linux Virtual Server (LVS)** una de las mejores opciones en software libre. Ha sido incorporado al núcleo del sistema operativo GNU/Linux y reproduce las capacidades de los switches capa 7 y pueden ser implantados a un costo significativamente inferior.

LVS puede trabajar bajo tres esquemas:

- Servidor NAT
- IP Tunneling
- Direct Routing

El primer esquema es sencillo pero no escalan bien, el segundo tiene una configuración compleja y el tercero lo consideramos el más apropiado, por tal razón nos enfocamos en éste último (figura 15: Balanceo de Carga con LVS). Para leer más sobre los dos primeros puede dirigirse al sitio web de LVS¹². Su funcionamiento es relativamente simple, un balanceador principal (BP) recibe todas las solicitudes y cambia la dirección IP de cada solicitud a la dirección IP del cliente correspondiente antes de encaminarla a uno de los servidores reales. Para reforzar la tolerancia a fallas se instala un balanceador secundario (BS) que prueba con una frecuencia determinada la disponibilidad del primero y en caso de

¹¹ Registro MX (Mail eXchanger) es uno de los registros del protocolo Domain Name Service (DNS) que se utiliza para responder a los clientes que solicitan los servidores de correo de un determinado dominio.

¹² <http://www.linuxvirtualserver.org/>





falla este cambia automáticamente su dirección IP por la del principal. La diferencia fundamental con la arquitectura de la figura 8: Configuración Estándar es que los nodos (servidores reales) están conectados directamente a la red pública para poder responder directamente a los clientes. De esta manera el balanceador de carga principal evita convertirse en un cuello de botella pues el tamaño de las solicitudes es relativamente pequeño y respecto al tamaño de las respuestas.

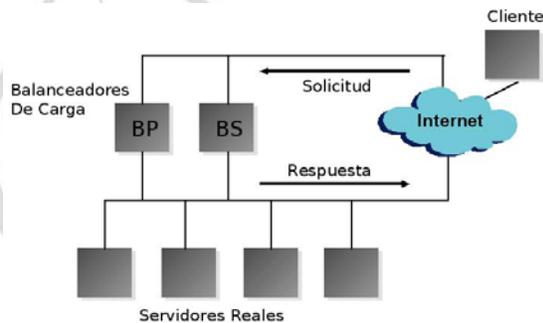


figura 15: Balanceo de Carga con LVS

5. Conclusiones

La redundancia natural que proveen los nodos de un cluster lo convierten en una plataforma confiable, tanto para la computación de alto rendimiento como para los servicios de redes. Así mismo, su alta escalabilidad, posibilidad de incorporación de nuevos recursos sin eliminar los existentes, inversión inicial y software de bajo costo convierten a los clusters Linux en la opción con el mejor factor precio-rendimiento del mercado. También podemos concluir que el desarrollo de aplicaciones de software libre y avances en tecnologías de procesadores, memoria, discos y redes para PCs han propiciado la difusión de esta tecnología en todos los campos donde se requiere el uso de plataformas de alto rendimiento. La experiencia lograda en el trabajo realizado en los últimos 10 años sobre clusters Linux tanto en Computación de Alto rendimiento como en plataformas adaptadas para Servicios de Redes nos hacen concluir adicionalmente que los clusters constituyen herramientas versátiles y de bajo costo para enfrentar altas demandas en servicios de distinta índole. En particular el escalamiento y la tolerancia a fallas que proporcionan los clusters son características fundamentales que permiten ofrecer servicios estables y de alto rendimiento

10. Referencias

[1] A.J. Forty (Chair), "Future Facilities for Advanced Research Computing" *The report of a Joint Working Party*, June 1985, Advisory Board for the Research Councils, Computer Board for Universities and Research Councils, University Grants Committee) SERC, 1985.

[2] R.G. Evans and S. Wilson (eds), *Supercomputational Science*, Plenum Press, New York, 1990.

[3] Spector, D. *Building LINUX Clusters*, O'Reilly, Sebastopol, California, 2002.

[4] Wilkinson, B., y Allen, M. *Parallel Programming: Techniques and Applications Using Networked Workstations and Parallel Computers*, Prentice Hall, Upper Saddle River, New Jersey, 1999.

[5] Gordon, B., y Gray J. *High Performance Computing: Crays, Clusters, and Centers. What Next?* Technical Report: MSR-TR-2001-76. Microsoft Corporation. 2001. <http://reserach.microsoft.com/pubs>

[5] Foster, I. *Designing and Building Parallel Programs: Concepts and Tools for Parallel Software Engineering*, Addison-Wesley, New York, 1995.

[6] Geist, A., Suderam, V., y otros. *PVM: Parallel Virtual Machine. A User's Guide and Tutorial for Networked Parallel Computing*, MIT Press, Massachusetts, 1994.

[7] Corner, D. *Computer Networks and Internet*. 2da. Edición. Prentice Hall, Upper Saddle River, New Jersey, 1999.

[8] Hoeger, H. *Introducción a la Computación Paralela*. Reporte Interno. Centro Nacional de Cálculo Científico Universidad de Los Andes. Mérida, Venezuela, 1997. http://www.cecalc.ula.ve/documentacion/manuales_tutoriales.html

[9] Snir, M., Dongarra, J., y otros. *MPI: The Complete Reference*. MIT Press, Massachusetts, 1996.

[10] Linux Rules Supercomputers, Daniel Lyons, 03.15.05, http://www.forbes.com/home/enterprisetech/2005/03/15/cz_d1_0315linux.html

[11] Moshe Bar, Stefano Cozzini, Mauricio Davini, Alberto Marmodoro *OpenMosix vx Beowulf: A case of study INFM Democritos (Italy)*, OpenMosix Project, Department of Physics University of Pisa (Italy)

10.1 Sitios en Internet

- 1) *CeCalCULA*:
<http://www.cecalc.ula.ve/>
- 2) Laboratorio Argon
<http://www.mcs.anl.gov/dbpp/>
- 3) Netlib





<http://www.netlib.org/pvm3/book/pvm-book.html/>

- 4) Información sobre *MPI: Message Passing Interface*: <http://www.mcs.an.gov/mpi/index.html>
- 5) Sitio con los 500 supercomputadores más poderosos: <http://www.top500.com>
- 6) National HPCC Software Exchange: <http://nhse.npac.syr.edu/>
- 7) Enciclopedia electrónica de términos computacionales: <http://www.webopedia.com>
- 8) Beowulf: <http://www.beowulf.org>

