

Servicios telemáticos sobre nubes privadas en plataformas virtualizadas y distribuidas

N. Arbós, L. M. Amorós, D. González, A. Oller, J. Alcober

Departamento de Ingeniería Telemática

Universitat Politècnica de Catalunya, Fundació i2cat

Esteve Terradas, 7 - 08860 Castelldefels.

noemi.arbos@i2cat.net, luismi.amoros@i2cat.net, david.gonzalez@i2cat.net, antoni.oller@upc.edu, jesus.alcober@upc.edu

Resumen – En la actualidad, el modelo para proporcionar servicios basados en alojar cada aplicación en un servidor físico no es el modelo óptimo, ya que los avances tecnológicos tanto en hardware como en software hacen que se produzca un desaprovechamiento de los recursos ofrecidos. Por este motivo, este trabajo presenta un nuevo modelo basado en un clúster de servidores con sistema de ficheros distribuido en red y virtualización de sistemas operativos que mejora la utilización de los recursos disponibles. Esta nueva arquitectura ofrece una nube privada que simplifica la gestión de los servicios y aporta una disminución del coste de mantenimiento, además de añadir flexibilidad, dinamismo y escalabilidad al sistema.

Palabras Clave- Virtualización, Clúster, NFS, Cloud Computing, IaaS.

I. INTRODUCCIÓN

Actualmente, la mayoría de entidades y empresas utilizan una gran cantidad de hardware, en especial servidores, para proporcionar sus servicios de infraestructura. Por cuestiones de seguridad y redundancia se ha ido siguiendo un modelo de un equipo para cada servicio. Sin embargo, el crecimiento exponencial de las capacidades de computación del hardware que ofrece el mercado [1], propicia que este modelo de utilización de los recursos no sea óptimo [2].

Por este motivo surge la virtualización, que trata de ofrecer una mejor utilización de los recursos de una máquina creando una capa de abstracción entre el hardware de la máquina física y el sistema operativo de la máquina virtual, el cual es totalmente independiente lo que permite que los posibles errores de funcionamiento de la máquina virtual no provoquen efectos colaterales a las demás.

Por otro lado, las arquitecturas de servicios evolucionan, y en este contexto, el paradigma de computación en nube, Cloud Computing [3], permite ofrecer un nuevo punto de vista para la construcción de servicios de computación sobre internet. Bajo este esquema y en una modalidad de nube privada se intenta potenciar el concepto de IaaS, Infrastructure as a Service, donde la arquitectura propuesta ofrece un nuevo modelo de servicio orientado especialmente a infraestructura.

El trabajo que se presenta, realiza una propuesta arquitectónica de nube privada basada en diferentes tecnologías, todas ellas independientes, reemplazables y adaptables a las diferentes herramientas existentes. Se trata de una plataforma virtualizada, dinámica y distribuida formada por dos tipos de nodos. Un primer tipo de nodos de alta

capacidad de procesamiento y un segundo tipo de nodos que proporciona el espacio de almacenamiento. Mediante un sistema de ficheros distribuido y utilizando protocolos de red, se integran ambos nodos. Este sistema distribuido mejora la escalabilidad y el tiempo total que el sistema está funcionando (Uptime) respecto a otras arquitecturas.

II. VIRTUALIZACIÓN

Virtualización es un término que se refiere a la abstracción de los recursos de una máquina. Esta capa de abstracción, más conocida como Hypervisor, se encarga de manejar y gestionar los recursos de hardware principales, CPU, disco, memoria RAM, etc. Por tanto, Hypervisor es el encargado de repartir estos recursos entre las diferentes máquinas virtuales que se estén ejecutando sobre la misma máquina física.

Dependiendo del hardware y del sistema operativo podemos distinguir hasta cuatro tipos de virtualización diferentes:

1) *Emulación o simulación*: el sistema operativo padre simula un hardware completo.

2) *Virtualización a nivel de Sistema Operativo*: se basa en crear celdas de usuarios independientes, sin acceso entre ellas.

3) *Paravirtualización*: la máquina virtual no simula un hardware, sino que ofrece una interfaz al sistema que sólo puede usarse mediante la modificación del kernel del sistema operativo de la máquina virtual.

4) *Virtualización nativa*: la máquina virtual simula un hardware subyacente completo para permitir un sistema operativo sin modificar en la máquina virtual.

Estas técnicas de virtualización ofrecen grandes mejoras y ventajas respecto a un sistema de hardware nativo:

- **Ahorro**: Mejora substancialmente el aprovechamiento de los recursos de hardware, lo que se traduce en un gran ahorro en cuanto a coste del hardware y también un ahorro en cuanto a energía consumida y espacio.
- **GreenIT [4]**: Al reducir el número de máquinas, reducimos la cantidad de energía utilizada, no sólo en alimentar el servidor sino que también se reduce el

consumo de refrigeración tanto eléctrico como las emisiones de aire caliente.

- **Flexibilidad:** En una misma máquina podemos tener funcionando juntos un sistema operativo Windows y un sistema basado en GNU/Linux [5].
- **Escalabilidad y Uptime:** La virtualización ofrece un sistema totalmente homogéneo a las máquinas virtuales.

III. ARQUITECTURA DEL SISTEMA

Tal como ilustra la figura 1, la arquitectura propuesta consta de 3 tipos de elementos:

- **Un Servidor Máster:** está conectado a una cabina de discos y su función es orquestar y controlar los servidores esclavos utilizando los recursos de la nube según sea necesario.
- **Clúster de Servidores Esclavos:** está formado por N servidores esclavos. Este tipo de nodos carecen de disco duro y, en consecuencia, son equipos limitados ya que sólo disponen de memoria RAM y CPU. Por lo tanto, los datos han de ser almacenados en la cabina de discos del servidor máster.
- **Máquinas Virtuales:** se crean dinámicamente sobre los nodos esclavos, aprovechando los recursos libres que el máster les asigna.

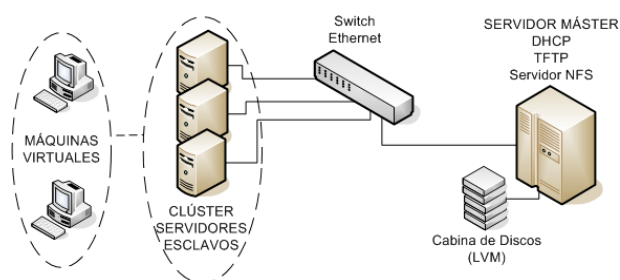


Fig. 1. Arquitectura del sistema.

Uno de los puntos críticos de la arquitectura es el proceso de inicialización. Los servidores esclavos cargan sus sistemas operativos a través de la red. Para solventar este problema se utiliza el protocolo PXE [G1], que consiste en la combinación de dos protocolos. El servicio DHCP [6], se utiliza para obtener una dirección IP y localizar el servidor de arranque. El servicio de TFTP [7], se utiliza para transferir un microkernel desde el servidor máster hasta el nodo esclavo. Una vez descargado el kernel se ejecuta en la memoria RAM. En nuestro caso, este kernel ha sido compilado con las opciones de virtualización y el sistema de ficheros esclavo ya dispone del software de virtualización Xen [8]. Se ha escogido Xen como sistema de virtualización ya que dispone de una versión open source muy eficiente y potente, que permite tanto paravirtualización como virtualización nativa, dependiendo de las necesidades, en contraste con otras plataformas como VirtualBox [9] o VMware [10].

Para conseguir dinamismo, el sistema de ficheros de los servidores esclavos se reserva en la cabina de discos mediante LVM [G2] y se monta utilizando el protocolo NFS [G3], lo que permite a las máquinas montar particiones en un sistema remoto y usarlas como si estuvieran en un sistema de archivos local. Este sistema nos permite ejecutar máquinas virtuales

sobre los servidores esclavos. Para facilitar la flexibilidad del sistema, estas máquinas virtuales montan el sistema operativo a través de NFS mediante LVM en la cabina de discos, de forma similar a los servidores esclavos.

Un ejemplo del funcionamiento del sistema lo podemos observar en la figura 2, donde encontramos un servidor esclavo (E.1) que ha arrancado el sistema operativo por red, mediante el protocolo PXE, y tiene el sistema de ficheros ubicado en la cabina de discos del servidor máster gracias al uso de NFS. También hay N máquinas virtuales ejecutándose en el esclavo (M.V.1, ..., M.V.N). Cada una de ellas hace uso de una proporción de RAM y CPU del servidor según sus necesidades, y tienen alojado el sistema de ficheros en la cabina de discos mediante NFS.

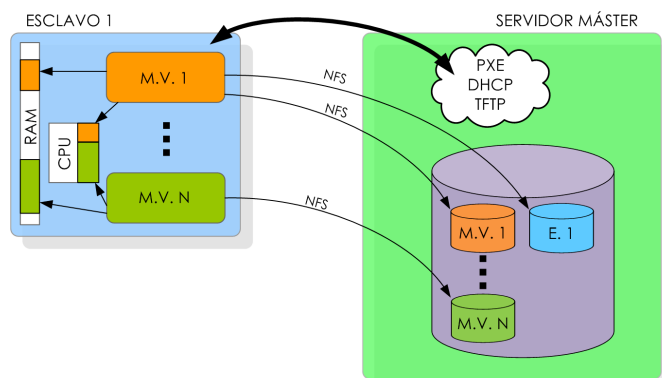


Fig. 2. Funcionamiento del sistema.

Este sistema nos ofrece flexibilidad debido a que, los ficheros/datos (tanto S.O como datos de usuario) están alojados en el máster y se puede orquestar y decidir donde se ejecutan las máquinas virtuales. Estas características nos permiten:

- **Live migration:** migración en tiempo real de las máquinas virtuales.
- Una gran tolerancia a fallos: si uno de los servidores esclavos falla, la plataforma, de manera dinámica, detecta la caída y se migran las máquinas virtuales a otro nodo.
- Balanceo de la carga: si un esclavo está saturado podemos migrar las máquinas a otro nodo.
- Escalabilidad: el número N de servidores esclavos no tiene cota, viene definido únicamente por las necesidades totales.
- Todas las ventajas que aporta la virtualización.

IV. ESCENARIO DE PRUEBAS

El escenario de pruebas consta de dos tipos de nodos con las siguientes características:

- **Servidor Máster:** utilizamos un equipo Supermicro X8DTN+ que dispone de un procesador Quad Core Intel Xeon E5504 Nehalem de 64 bits, 12 GB de memoria RAM DDR3 y una cabina de discos de 15 TB.
- **Servidores Esclavos:** utilizamos 10 equipos Supermicro 5016T-MRB que dispone cada uno de un procesador Quad Core Intel Xeon E5504 Nehalem de 64 bits y 12 GB de memoria RAM DDR3.

El escenario, tal como muestra la figura 3, está formado por 2 segmentos de red. En el primero utilizaremos direccionamiento privado y se utilizará para el PXE y el NFS de los servidores esclavos, en el segundo segmento utilizaremos direccionamiento público para acceder a Internet y el NFS de las máquinas virtuales.

Por lo tanto, todos los servidores han de disponer de dos tarjetas de red, una para la red privada y otra para la red pública. Las máquinas virtuales, sin embargo, sólo tendrán acceso a la red pública.

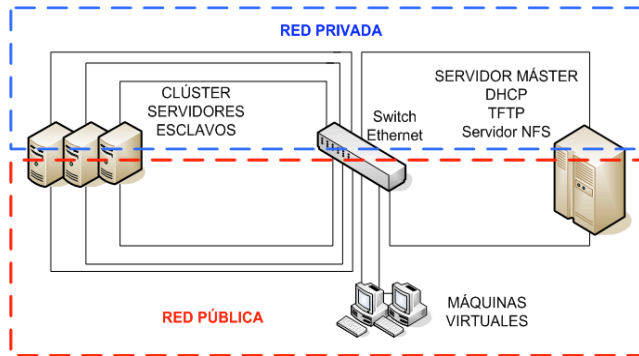


Fig. 3. Escenario de pruebas.

El servicio de NFS está configurado para funcionar utilizando TCP, que asegura que llegan todos los datos de un punto a otro sin errores, con permisos de escritura y lectura para el usuario remoto. También se ha configurado de tal manera que los usuarios remotos pueden actuar como *root* en el sistema local. Como método de escritura, se ha escogido la opción de escritura síncrona, es decir, todas la escrituras en el disco se realizan antes de devolver el control al cliente remoto, aunque esta opción puede disminuir el rendimiento se asegura que no se pueden perder datos si el servidor principal cae o se apaga.

V. PRUEBAS DE RENDIMIENTO

El modelo de innovación de esta arquitectura se basa en la flexibilidad que ofrece el sistema de discos por red y la nube de servidores esclavos.

Este sistema tiene un inconveniente principal, la pérdida de rendimiento respecto a un sistema en local debido al uso del protocolo NFS. Por esta razón, se han realizado pruebas tanto de lectura como escritura en disco comprobando el rendimiento de los servidores esclavos y las máquinas virtuales a través de NFS y de esta forma validar el sistema. Todas las pruebas han sido realizadas con la herramienta Spew [11].

Pruebas de escritura en disco: consisten en la escritura de un megabyte de datos utilizando bloques de 512 bytes en un fichero determinado. Se han realizado 50 iteraciones de este proceso y se ha extraído la media aritmética de las tasas de escritura.

WTR (Mbps)

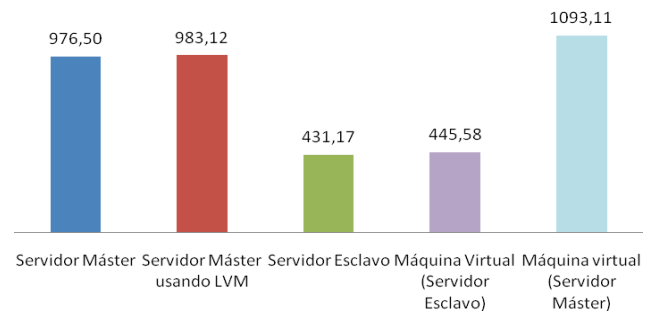


Fig. 4. Resultados de Write Transfer Rate.

Según los resultados obtenidos en las pruebas de escritura en disco, tener el sistema de ficheros de los servidores esclavos y las máquinas virtuales mediante NFS causa que la tasa de escritura disminuya entre un 50-55% respecto al servidor principal.

La causa de esta pérdida de rendimiento del NFS en escritura se debe a que utilizamos escritura síncrona, por lo tanto, se realizan todas las operaciones de escritura en disco solicitadas antes de devolver el control al cliente. De esta forma, disminuye el rendimiento obtenido, pero por otra parte, nos aseguramos que los datos se guardan correctamente

Además, el hecho de utilizar TCP puede provocar retransmisiones en caso de errores en la transmisión. Este hecho podría causar una disminución del rendimiento de nuestro sistema NFS, tanto en lectura como en escritura. Es cierto que se podría haber utilizado UDP ya que se trata de una red de área local donde la tasa de error es realmente baja, pero un pequeño error en el envío de información podría haber causado la inutilización del servicio alojado en las máquinas virtuales.

Pruebas de lectura en disco: consisten en la lectura de un megabyte de datos utilizando peticiones de 512 bytes de un fichero determinado. Se han realizado 50 iteraciones de este proceso y se ha extraído la media aritmética de las tasas de lectura.

RTR (Mbps)

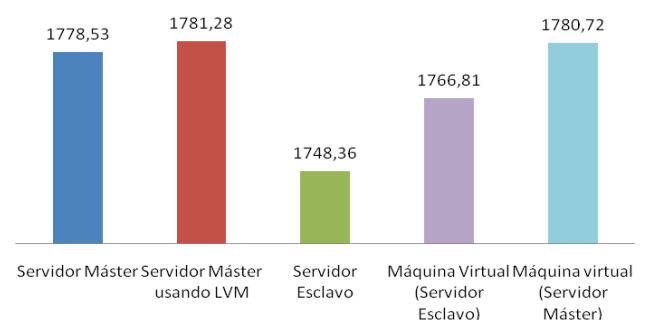


Fig. 5. Resultados de Read Transfer Rate.

Las pruebas realizadas en lectura de disco han sido satisfactorias, ya que la pérdida de tasa de lectura respecto al servidor principal ha sido de entre un 1-2%, en el caso de un servidor esclavo; y de entre un 0,5-1% en las máquinas virtuales ejecutadas en un servidor esclavo, lo que nos indica un rendimiento muy alto en lectura.

VI. APLICACIONES Y BENEFICIOS

Una de las principales ventajas es la mejor utilización del rendimiento de un servidor físico para crear más de un servidor virtualizado, dedicando una máquina virtual para cada aplicación. Esta alternativa permite utilizar la mayor parte de esos recursos creando servidores virtuales aislados entre sí de forma que existen varios servicios con una única máquina física.

La arquitectura propuesta, basada en un servidor físico que aloja varios servidores virtuales, además de extraer un mayor rendimiento de los recursos, es una solución escalable, barata y fácil de mantener. El hecho de utilizar máquinas virtuales hace posible trasladarlas de una máquina física a otra en caso de fallo de una de éstas, mediante *Live migration*: mover máquinas virtuales entre servidores físicos sin detener la prestación de servicios, proporcionando una gran flexibilidad. Además, el mantenimiento del hardware y el coste energético disminuye respecto a la arquitectura un servidor físico sobre un servicio, ya que utilizamos menos máquinas físicas para muchos más servicios.

Otro de los beneficios que aporta esta nueva arquitectura, es que los sistemas operativos virtualizados no tienen dependencia del hardware, ya que se crea una capa de abstracción entre el sistema operativo utilizado y el hardware real. De esta forma, podemos tener máquinas virtuales con sistemas operativos diferentes en una misma máquina física.

Desde el punto de vista del usuario final, el hecho de que los servidores físicos monten su sistema de ficheros principal por NFS y arranquen mediante PXE, es totalmente transparente. Lo mismo ocurre con el hecho de utilizar máquinas virtuales para alojar los servicios, ya que cada servidor virtual tendrá las mismas prestaciones que un servidor físico tradicional.

VII. CONCLUSIONES

En este artículo se ha presentado un nuevo sistema con unas características que hacen que sea una opción eficiente frente a otro tipo de arquitecturas, esto es debido al modelo de arquitectura presentado en este trabajo y las características y dinamismo que ofrece.

Por otra parte, se ha comprobado que el rendimiento final que se puede obtener es satisfactorio ya que a pesar que las pérdidas que genera la arquitectura pueden llegar a suponer un coste de casi un 50% en la tasa de escritura, la flexibilidad que ofrece esta arquitectura compensa con creces estas pérdidas en entornos donde la escritura en disco no supone un requerimiento muy importante. En cambio, los resultados de las tasas de lectura en disco son muy buenos, ya que se obtienen valores similares a las tasas de lectura en un entorno de acceso físico al espacio de almacenamiento. Esto es especialmente importante en entornos de servicios, en especial web, donde se reciben miles de peticiones de lectura por segundo y prima más el tiempo de respuesta ante fallos y la capacidad de escalar el sistema que la tasa de escritura en disco.

Los diferentes protocolos utilizados, como PXE, NFS, TFTP, etc., y la virtualización son totalmente transparentes al usuario final, de esta manera, se consiguen todas las ventajas de utilizar esta nueva arquitectura sin afectar a la visión del

usuario final del sistema que acceda a una de los servicios ofrecidos.

Esta arquitectura puede ser ampliada para hacer más sencilla la tarea de administración del clúster de servidores y máquinas virtuales mediante un sistema de gestión o controlador. También se puede mejorar la creación y monitorización de las máquinas virtuales a través de una herramienta de gestión específica, como por ejemplo OpenNebula [12] o Ganeti [13].

De esta manera, se ofrece una plataforma distribuida con un paradigma basado en nube potenciando la infraestructura como servicio, y ofreciendo una arquitectura sobre la que desplegar servicios.

AGRADECIMIENTOS

Este trabajo ha sido apoyado por el MCyT (Ministerio de Ciencia y Tecnología del Gobierno de España) en el marco del proyecto TSI2007-66637-C02-01, parcialmente financiado por el FEDER.

REFERENCIAS

- [1] Ley de Moore
[http://es.wikipedia.org/wiki/Ley_de_Moore]
- [2] J. Pujal; A. Oller; J. López; C. Fanning; F. Minerva; J. Alcober; Escritorios remotos en máquinas virtuales aplicados en grandes corporaciones; Boletín de RedIRIS, nº 85-86, marzo-2009 Pág. 42-49.
[<http://www.rediris.es/difusion/publicaciones/boletin/85-86/ponencias85-5.pdf>]
- [3] M. Armbrust; A. Fox; R. Griffith; A. Joseph; R. Katz; A. Konwinski; G. Lee; D. Patterson; A. Rabkin; I. Stoica; M. Zaharia; Above the Clouds: A Berkeley View of Cloud Computing; UCB/EECS; Technical Report No. UCB/EECS-2009-28, febrero-2009.
[<http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html>]
- [4] GreenIT – sustainable information technology
[<http://www.greenit.net/>]
- [5] Linux
[<http://www.linux.org/>]
- [6] R. Droms; RFC 2131 “Dynamic Host Configuration Protocol”; marzo-1997.
[<http://www.rfc-editor.org/rfc/rfc2131.txt>]
- [7] K. Sollins; RFC 1350 “Trivial File Transfer Protocol”; julio-1992.
[<http://www.rfc-editor.org/rfc/rfc1350.txt>]
- [8] Xen
[<http://www.xen.org/>]
- [9] VirtualBox
[<http://www.virtualbox.org/>]
- [10] VMware
[<http://www.vmware.com/es/>]
- [11] Spew
[<http://spew.berlios.de/>]
- [12] OpenNebula
[<http://opennebula.org/>]
- [13] Ganeti
[<http://code.google.com/p/ganeti/>]
- [14] D. González; A. Oller; Desarrollo de una plataforma de virtualización; Bachelor thesis, UPC, EPSC, 3-abr-2008.
[<http://upcommons.upc.edu/pfc/handle/2099.1/4798>]
- [15] J. Pujal; Proposal of remote virtual desktops architecture working with thin client devices; Master thesis, UPC, 16-jul-2008.
[<http://upcommons.upc.edu/pfc/handle/2099.1/5320>]

GLOSARIO

- [G1] PXE: Preboot eXecution Environment
[<http://es.wikipedia.org/wiki/PXE>]
- [G2] LVM: Logical Volume Manager
[<http://es.wikipedia.org/wiki/LVM>]
- [G3] NFS: Network File System
[http://es.wikipedia.org/wiki/Network_File_System]