

# Análisis de mejora en la escalabilidad de las infraestructuras de cloud computing

Nelson Rodríguez, María Murazzo, Diego Medel, Susana Chavez, Adriana Martín, Adriana Valenzuela

Departamento e Instituto de Informática  
Facultad de Ciencias Exactas, Físicas y Naturales – UNSJ  
Complejo Islas Malvinas (CUIM), San Juan, Argentina

## Resumen

En los últimos años y con el auge de las tecnologías de virtualización y de las infraestructuras cloud, se abre un nuevo abanico de posibilidades para el acceso a recursos de cómputo para el ámbito científico. Las tecnologías cloud permiten “acceso ubicuo, adaptado y bajo demanda en red a un conjunto compartido de recursos de computación”. Estas tecnologías permiten que el acceso a grandes cantidades de recursos virtualizados sea mucho más sencillo.

Cloud computing provee un ambiente de aprovisionamiento de recursos en Internet, bajo demanda basado en la virtualización. En la actualidad existe un creciente interés en usar esta tecnología para resolver problemas científicos e industriales de gran envergadura.

Lo más atractivo de la migración a una infraestructura cloud es el bajo costo de la solución adoptada, pues no es necesario adquirir equipamiento específico y costoso. Sin embargo un aspecto a tener en cuenta es la necesidad de evaluar el costo computacional de la migración debido a la pérdida de performance ocasionada por los retardos en las comunicaciones y en la virtualización.

El presente trabajo presenta una línea de investigación cuyo objetivo será la evaluación de la performance de la virtualización en infraestructuras de cloud computing.

**Palabras Claves:** cloud computing, IaaS, open source, migración, virtualización.

## 1. Introducción

El auge en las técnicas de virtualización durante estos últimos años ha propulsado la aparición del Cloud Computing, ofreciendo el acceso bajo demanda y a través de Internet a un conjunto de recursos virtualizados configurables, y que cuyo aprovisionamiento y repliegue se realiza sin mayor esfuerzo del usuario.

En función de ello, Cloud Computing se ha convertido en un enorme repositorio de recursos computacionales, lo cual es una buena posibilidad para construir una plataforma para las aplicaciones que necesitan una gran cantidad de recursos. Esta capacidad del cloud se debe principalmente a la habilidad de escalado elástico de recursos en función de las necesidades de las aplicaciones y el presupuesto del usuario. Esta es una

tecnología centrada en ofrecer cómputo bajo demanda como cualquier otro servicio. Otro aspecto que afecta la performance de las aplicaciones al montarlas sobre arquitectura cloud, es la capacidad de virtualización de cualquier recurso (bases de datos, red, procesador, etc.) y ofrecerlo como un servicio (AaaS, Anything as a Service) [1].

Sin embargo y a pesar de estas ventajas, el principal problema al que se debe hacer frente es la escasa performance en el acceso a Internet, la cual puede generar retardos, que impactaran de manera negativa en la ejecución de cierto tipo de aplicaciones (por ejemplo, las de cálculo intensivo con restricciones de calidad, donde es necesario mucha transferencia de información).

Para solucionar este problema una posibilidad es montar una infraestructura cloud privada con el objetivo de sacar el máximo provecho a las bondades de la virtualización de recursos, sin pagar el precio de la degradación de la calidad por tener que acceder a Internet.

## 2. Arquitectura de Cloud Computing

Cloud Computing es un modelo para permitir el acceso en red omnipresente, conveniente y bajo demanda a un conjunto compartido de recursos informáticos configurables tales como, redes, servidores, almacenamiento, aplicaciones y servicios que pueden ser rápidamente aprovisionados y liberados con una mínima gestión esfuerzo o interacción con el proveedor o administrador de servicio. Por ejemplo, en lugar de comprar un servidor, switches, firewalls, cables, etc. y alojarlos físicamente, se puede contratar capacidad de procesamiento, memoria, disco y ancho de banda a un proveedor de cloud hosting; esto es Infraestructura Cloud.

El NIST (National Institute of Standards and Technology), ha presentado una de las definiciones de Cloud más clara y comprensible. La define como “*un modelo tecnológico que permite el acceso ubicuo, adaptado y bajo demanda en red a un conjunto compartido de recursos de computación configurables (por ejemplo: redes, servidores, equipos de almacenamiento, aplicaciones y servicios), que pueden ser rápidamente aprovisionados y liberados con un esfuerzo de gestión reducido o interacción mínima con el proveedor del servicio*” [2].

Según dicha definición, Cloud Computing se entiende como un modelo de prestación de servicios informáticos cuya principal orientación es la escalabilidad. Desde el punto de vista de los usuarios, los servicios son elásticos, es decir, que pueden crecer o recuperar su tamaño original de manera rápida y sencilla. Esta orientación permite que los usuarios que acceden a los servicios, perciban que todo funciona de manera simple y rápida, dando como resultado una experiencia más gratificante.

Para lograr esto, cloud esta organizada en una arquitectura de capas llamada Modelo de Servicio, donde el *SaaS (Software as a Service)* se encuentra en el tope del stack, es el nivel más visible y permite a los usuarios tener acceso a las aplicaciones alojadas en el cloud (por ejemplo: Google App). El *PaaS (Plataforma as a Service)* es una capa de abstracción o middleware que se encuentra entre el SaaS y la infraestructura virtualizada, y su objetivo es proveer a los desarrolladores un framework y un conjunto de API's que puedan usar para construir sus aplicaciones (por ejemplo: Google App Engine). Por ultimo, el *IaaS (Infraestructura as a Service)* ofrece una infraestructura virtualizada de recursos (procesador, red, base de dato, disco, etc.) como servicio, de forma de ofrecer a los usuarios un ambiente con recursos escalables de manera elástica y pagando por su uso.

Un aspecto muy importante de esta arquitectura es la capa de *Virtualización* que se ubica como un middleware sobre el IaaS y es la encargada del aislamiento y la flexibilidad en el manejo de recursos. Gracias a esta capa de virtualización, es posible la compartición de recursos permitiendo la asignación, reasignación y desasignación de “instancias” de cada recurso. Para lograr esto es necesario contar con mecanismos de “reservación de recursos” y de “administración de recursos” para permitir el manejo consistente de todas las instancias de los recursos virtualizados con el objeto de lograr su máxima performance.

Por otro lado, se encuentra el Modelo de Despliegue, que se refieren a la posición (localización) y administración (gestión) de la infraestructura de cloud y pueden ser *públicos* (manejadas por terceras partes), *comunitarios* (operada por varias organizaciones para uso compartido) o *privados* (propiedad de una organización), como una cuarta opción aparece el modelo *híbrido* como la combinación de cualquiera de los tres modelos anteriores.

Hay que aclarar, que cuando se decide contar con una solución basada en cloud es de suma importancia analizar cual despliegue se usará, pues esta elección impacta directamente en la performance de la solución. Hay que recordar que cualquier elección que implique el uso intensivo de comunicaciones degrada el desempeño por los retardos incurridos en el proceso de comunicación en si [3].

Cloud Computing provee una plataforma en la cual los usuarios interactúan con aplicaciones de forma remota sobre Internet, bajo la premisa de la compartición de recursos; permitiendo acceder a una plataforma donde se provee “cualquier cosa”: poder de cómputo, infraestructura, almacenamiento, comunicaciones y servicios en cualquier lugar y en cualquier momento.

Cloud Computing tiene cuatro características principales: elasticidad, escalabilidad, aprovisionamiento dinámico de recursos y pago por el uso en función de la elasticidad en el consumo. Esta flexibilidad en la administración y optimización de recursos es uno de los aspectos más atractivos para que la migración al Cloud Computing se concrete.

### 3. Migración al Cloud Computing

La constante demanda de mayor poder computacional, no sólo de la comunidad científica y académica, ha determinado grandes avances en el hardware de computadoras. La computación de alto rendimiento (HPC, High Performance Computing) se encarga de explotar los recursos de tecnologías computacionales como son los clúster, supercomputadoras o mediante el uso de la computación paralela, para obtener resultados y/o soluciones a simulaciones que requieren gran capacidad de cómputo.

Si bien la mayor ventaja de la HPC es la capacidad de hacer gran cantidad de cómputo en poco tiempo, su mayor desventaja es la necesidad de contar con hardware caro de instalar, mantener y operar. Otro problema de las plataformas de HPC tradicionales es el acceso a hardware limitado y las dificultades de escalado en función de los requerimientos de las aplicaciones. Este ultimo problema, es la principal razón por la que se esta comenzando a migrar al Cloud Computing como plataforma para realizar computo de de alto rendimiento.

La migración hacia el Cloud Computing ha sido impulsada en gran medida por el bajo costo que representan. Sin embargo, el menor costo no siempre es una buena razón debido a que el usuario puede tener otros requerimientos en términos de fiabilidad y disponibilidad del servicio.

En función de esto, se puede decir que la principal motivación de mover aplicaciones al cloud es reducir los costos de mantenimiento y abastecimiento, al tiempo que se aumenta la escalabilidad y la disponibilidad. Cloud Computing ofrece una forma de eliminar la necesidad de mantener el hardware caro permitiendo a los usuarios y desarrolladores centrarse en las aplicaciones.

Un aspecto muy importante a tener en cuenta es que, si bien migrar al cloud tiene muchas ventajas, también tiene desventajas, entre las cuales se puede mencionar como la más relevante las demoras incurridas en usar una arquitectura distribuida donde no existe un

sincronismo global y las comunicaciones degradan la performance de la infraestructura.

En función de esto, se debe tratar de minimizar los efectos provocados por la migración a un entorno distribuido mediante la elección adecuada de herramientas que permitan minimizar los impactos negativos sobre la performance de las aplicaciones.

### 3.1. Parámetros para la migración

Una de las características más relevantes del Cloud Computing es su elasticidad. Los entornos Cloud son escalables, es decir, capaces de ajustarse a la demanda de los usuarios. Ante un pico elevado de requisitos de cómputo, un usuario puede solicitar mayor capacidad de cálculo, que provocará el despliegue automático de nuevas instancias sobre la infraestructura física existente.

Por el contrario, si lo que se detecta es una baja de la demanda de cómputo, el usuario puede decidir prescindir de parte de su infraestructura, pudiéndola ajustar a su uso de forma dinámica. Esta es una gran ventaja, puesto que uno de los problemas que existían hasta la aparición del Cloud Computing era la creación de una infraestructura ad hoc preparada para soportar demandas no previsibles de cómputo, ya que esto suponía una elevada inversión y además no se podía asegurar que fuese capaz de responder correctamente a estos incrementos repentinos en la demanda de los recursos de cómputo.

Desde el punto de vista de la arquitectura de Cloud Computing la capa más involucrada en la migración hacia el cloud es el IaaS, pues es la que dará el soporte de hardware (físico o virtualizado) para concretar la migración.

El IaaS brinda la capacidad de cómputo, procesamiento y almacenamiento para que el usuario pueda desplegar de forma transparente sistemas operativos y aplicaciones. Para lograr esta transparencia la infraestructura se monta sobre una capa de virtualización que permite contar con un ambiente optimizado independientemente del hardware físico mediante la generación de instancias abstractas de recursos llamadas máquinas virtuales, las cuales se encuentran administradas por los hipervisores [4].

Para lograr esto, los productos disponibles en el mercado para instalar infraestructuras cloud cuentan con diferentes hipervisores encargados de orquestar los recursos virtuales de cada máquina virtual con el objeto de sacar el máximo beneficio a los recursos físicos.

En función de esto, aparecen dos problemas que deben ser paliados con el objetivo que la migración sea exitosa. En primer lugar, hay que tratar de *minimizar los retardos* incurridos por las comunicaciones. Si este aspecto no es tratado de forma adecuada, puede degradar

la performance de las soluciones que tengan fuertes restricciones de tiempo; tales como el cómputo en tiempo real. Una manera de disminuir el impacto de este problema es contar con una infraestructura cloud privada, de esta manera se evita depender del cloud público y así eliminar el costo de la conexión hacia él.

En segundo lugar, hay que tratar de *maximizar la escalabilidad* de la infraestructura en la medida que exista un aumento en la demanda de recursos virtualizados. Esta escalabilidad se mide en función de tres parámetros: overhead, linealidad y aislamiento:

- El overhead es el tiempo de procesamiento requerido por una máquina virtual antes de la ejecución de un comando.
- La linealidad es una evaluación cuantitativa de cuán fuertemente relacionado está un conjunto de datos.
- El aislamiento asegura que en una situación de desequilibrio de la carga entre máquinas virtuales, todas las máquinas virtuales obtendrán recursos de manera equitativa.

La escalabilidad de una infraestructura cloud está fuertemente ligada con la tecnología de virtualización o hipervisores que se está usando. Por esta razón es tan importante la elección adecuada del hipervisor. Hay que destacar que si bien la virtualización aísla al usuario / desarrollador de los detalles de orquestación de recursos, un aspecto que se debe analizar es el overhead producido por los hipervisores. Estos, consumen recursos para la administración de los recursos virtualizados, lo que lleva a que mientras más se virtualice, más overhead haya y por consiguiente más sobrecarga y menos rendimiento [5].

Cabe destacar que cuando se trabaja sobre una plataforma HPC, ya sea con memoria compartida (multicore) o memoria distribuida (cluster), se debe analizar el problema de las comunicaciones. En el primer caso las comunicaciones no poseen impacto directo en el rendimiento de los algoritmos que se están ejecutando a menos que se deba acceder a memoria principal. Sin embargo en el caso de los cluster el problema de las comunicaciones implica una degradación importante en la ejecución pues se deben usar una solución basada en pasaje de mensajes.

Pero cuando se está trabajando sobre una infraestructura cloud, se agrega el problema de la administración de la virtualización, el cual incluye:

- *Tiempo de acceso al catálogo*: todos los IaaS poseen un catálogo de hipervisores que permite seleccionar la imagen más apropiada en base a los requisitos de la aplicación. Pero el acceso a este servicio y su posterior obtención de los resultados, que puede involucrar la descarga de la imagen, suponen un tiempo que es necesario tener en cuenta a la hora de decidir qué infraestructura es la

recomendada. La transferencia de la imagen dependerá de la calidad y el estado de la red en el momento del envío.

- *Tiempo de despliegue de los hipervisores:* otro parámetro concreto de este entorno hace referencia al tiempo consumido para poner en marcha un hipervisor. Este periodo dependerá del IaaS que se use (OpenStack, OpenNebula, Eucalyptus, Nimbus,...), además del estado en el que se encuentre la infraestructura donde se van a desplegar los hipervisores. Es más, si se utilizan técnicas de Green Computing, el despliegue del hipervisor puede provocar el encendido de los nodos, tiempo que habrá que contabilizar en este parámetro.
- *Tiempo de contextualización:* para conseguir ejecutar con éxito el trabajo del usuario en la instancia del hipervisor, será necesario aplicar un proceso de contextualización basado en la ejecución de scripts para dotarla de los requerimientos de software necesarios y preparar el entorno de ejecución. Además, este proceso se encargará de dejar instalada la aplicación para poder ser posteriormente ejecutada. La instalación típicamente consistirá en la descompresión y compilación del trabajo.
- *Existencia de VMs previamente desplegadas:* un usuario puede disponer de hipervisores ya desplegadas y contextualizadas, por ejemplo, debido a una ejecución previa. En este caso, la sobrecarga que supone el empleo de la infraestructura Cloud se vería considerablemente reducida.

En función de todo esto es necesario realizar un análisis de estos parámetros a la hora de seleccionar la infraestructura a usar.

#### 4. Infraestructura Propuesta

Con el objetivo de medir el desempeño de diferentes hipervisores, se ha decidido realizar la instalación de un cloud privado open source. Para ello, el mercado ofrece muchos productos de los cuales se ha realizado un estudio comparativo de tres: OpenStack, OpenNebula y CloudStack.

El parámetro que se ha tomado para realizar la evaluación son los hipervisores soportados y si permiten administrar virtualización por hardware.

- *CloudStack:* software open source para crear, controlar y desplegar infraestructuras cloud. Originalmente fue desarrollado por cloud.com y posteriormente por Citrix, actualmente es un proyecto de la Apache Software Foundation (ASF). Implementa Amazon EC2, S3 APIs, vCloud API y

su propia API CloudStack, escrita en Java. Soporta Xen, KVM, VMware y acceso a Amazon EC2s.

- *OpenNebula:* conjunto de herramientas open source que permiten manejar la infraestructura de virtualización de un centro de datos para construir cloud privadas, públicas e híbridas. Esta escrita en C++, Ruby y Shell. Soporta Xen Server, KVM, vSphere, y acceso a Amazon EC2s.
- *OpenStack:* Es una herramienta de colaboración global, libre y de open source distribuido bajo los términos de la licencia Apache. Es una de las iniciativas de cloud que cuenta con más apoyos, entre los que están Dell, Cisco y HP, así como RackSpace. Esta escrita en Python y soporta ESX, Hyper-V, KVM, LXC, QEMU, UML, Xen y XenServer.

De estos tres productos, se ha seleccionado OpenStack, debido a que es el proyecto que mayor aceptación y mayores contribuciones ha tenido en los últimos años [6]. Además, es el producto con mayor número de hipervisores soportados, lo que lo hace adecuado para poder realizar una evaluación más amplia de su desempeño.

OpenStack [7] es una iniciativa que busca la interacción estandarizada de las plataformas cloud de diferentes empresas. Es un proyecto global para crear una plataforma de cómputo open source en el cloud, que cumpla con las necesidades de los proveedores de servicios en clouds públicos, privados e híbridos, independientemente de su tamaño, que sea fácil de implementar y escalable.

El proyecto fue fundado en octubre de 2010 por la empresa Rackspace Cloud y por la agencia espacial estadounidense, NASA. Actualmente es gestionado por la Fundación OpenStack y cuenta con el apoyo de más de 178 empresas, entre las que destacan Rackspace, Oracle, AMD, Cisco, Canonical, IBM, HP, Dell Red Hat, Suse Linux, VMware, Yahoo y KIO Networks en América Latina.

La arquitectura de OpenStack es totalmente modular y en la actualidad engloba dos proyectos relacionados: *OpenStack Compute* y *OpenStack Object Storage*, tal como se muestra en la figura 1. El primero, denominado NOVA, aportado por la NASA, consiste en software dedicado al aprovisionamiento y administración de un gran número de servidores privados virtuales. El segundo proyecto, denominado SWIFT, aportado por Rackspace, permite la creación de almacenamiento redundante, escalable y de alta disponibilidad e integridad para terabytes o petabytes de datos.

Además, dependiendo de la versión de la que se trate se cuentan con otros componentes [8]. Cada uno de estos componentes se comunican mediante APIs provistas por el componente.

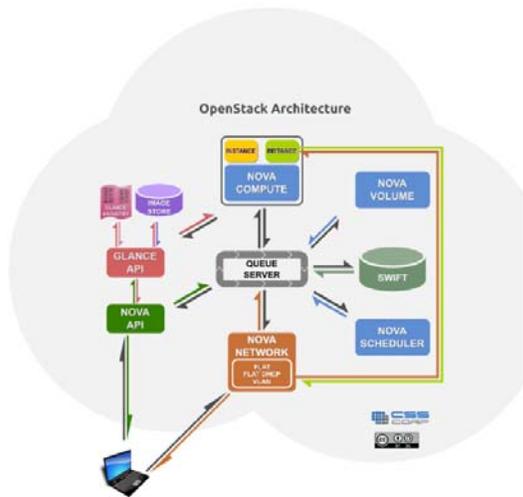


Figura 1: Arquitectura de OpenStack

En la actualidad se cuenta con una infraestructura instalada OpenStack Juno [9]. Esta instalación se ha realizado con una arquitectura monolítica, esto es que todos los componentes se han montado en un solo nodo. Esta arquitectura tiene la ventaja que elimina los retardos de la comunicación entre los componentes pues las comunicaciones se realizan en forma local, pero el inconveniente que posee es que solo se puede usar el hipervisor QEMU [10], lo cual elimina la posibilidad de evaluación de performance con otros hipervisores.

## 5. Trabajos Futuros y Conclusiones

Para poder realizar la evaluación de performance de los hipervisores se ha decidido migrar a una arquitectura distribuida, esta se montará sobre tres nodos y brindará la posibilidad de seleccionar con que hipervisor levantar las máquinas virtuales. Incluso permite que en una misma sesión se levanten varias máquinas virtuales con diferentes hipervisores.

El mayor inconveniente de esta arquitectura es que se va a tener que evaluar el impacto de las comunicaciones en red, pues los nodos estarán conectados entre sí.

Otra migración que se prevee es la migración a la última versión estable, que se liberó en abril de este año, denominada Kilo [11]. Esta versión tiene la ventaja de poder trabajar con un componente nuevo, denominado IRONIC que provee virtualización nativa, de esta manera el proceso de administración de las instancias será más rápido y hará uso más eficiente de los recursos físicos.

Hasta el momento, la conclusión más importante a la que se ha llegado es que OpenStack brinda un ambiente de infraestructura robusto, escalable y estable, debido a un sistema de versionado que cada seis meses libera una

versión con mejoras a los componentes existentes y con nuevos componentes. Además, otra ventaja importante es la provisión por parte de OpenStack de herramientas que facilitan el proceso de instalación y configuración de los componentes.

## 6. Bibliografía

- [1] Schubert, Jeffery, Neidecker-Lutz. "The Future of Cloud Computing: Opportunities for European Cloud Computing Beyond 2010:—expert Group Report. European Commission, Information Society and Media, 2010.
- [2] Mell, Grance. "The NIST definition of cloud computing." NIST Special Publication 800 – 145, 2011.
- [3] Rodríguez, Chávez, Martín, Murazzo, Valenzuela. "Interoperabilidad en cloud computing". WICC 2011. Rosario Argentina.
- [4] Rahma, Adji, Widyawan. (2013). "Analysis of KVM-Based Private Cloud for IaaS. *International Journal of Cloud Computing and Services Science*", (IJ-CLOSER), 2(4), 288-295. 2013
- [5] Huber, von Quast, Hauck, Kounev. "Evaluating and Modeling Virtualization Performance Overhead for Cloud Environments". In *CLOSER* (pp. 563-573). 2011
- [6] Jiang. "Open source IaaS community analysis. OpenStack vs OpenNebula vs Eucalyptus vs CloudStack". School of Information Technologies. 2015
- [7] OpenStack. "OpenStack: The Open Source Cloud Operating System". URL: <http://www.openstack.org/software/>
- [8] Rodríguez, Murazzo, Chávez, Guevara. "Arquitectura de cloud computing híbrida basada en tecnología open source". *XX Congreso Argentino de Ciencias de la Computación. 2014*.
- [9] OpenStack. "OpenStack Juno". URL: <https://www.openstack.org/software/juno/>
- [10] QEMU: "QEMU, Open Source Processor emulator". URL: <http://www.qemu.org>
- [11] OpenStack. "OpenStack Kilo". URL: <https://www.openstack.org/software/kilo/>