

Evaluación de arquitecturas distribuidas de commodity basadas en software libre

Nelson Rodríguez¹, María Murazzo², Adriana Martín³, Susana Chávez⁴, Diego Medel⁵, Jorge Mercado⁶

¹⁻⁵Departamento e Instituto de Informática - F.C.E.F. y N. - U.N.S.J.

⁶Departamento de Matemática - Facultad de Ingeniería - UNSJ

Complejo Islas Malvinas. Cereceto y Meglioli. 5400. Rivadavia. San Juan, 0264 4234129

¹nelson@iinfo.unsj.edu.ar ²marite@unsj-cuim.edu.ar ³arianamartinsj@gmail.com

⁴schavez@iinfo.unsj.edu.ar ⁵vdiego.unsj@hotmail.com ⁶jorgenmp@gmail.com

Resumen

La evolución de las tecnologías asociadas a la conectividad y al procesamiento distribuido ha permitido que surjan nuevas soluciones y necesidades por parte de los usuarios. Nuevos modelos de negocios están emergiendo gracias a la infraestructura de tecnologías informáticas distribuidas como Cluster y Cloud. La variedad de arquitecturas distribuidas ofrecen distintas prestaciones para centros de datos pequeños, redes locales de empresa, Clouds privados o plataformas distribuidas de cómputo. Esta línea de investigación plantea un análisis profundo de todas estas variantes de arquitecturas distribuidas de bajo costo. A diferencia del análisis propuesto, las empresas proveedoras de servicios Clouds ofrecen soluciones adecuadas a costa de una infraestructura sobredimensionada en software y hardware. Diversas organizaciones pequeñas o medianas necesitan procesar sus datos en sus propios servidores. Estas organizaciones no cuentan con recursos ilimitados y la eficiencia de los mismos es el problema que trata de resolver el presente proyecto.

Palabras clave: *Distributed Computing, Cloud Computing, Cluster, SaaS, PaaS*

Contexto

El presente trabajo se encuadra dentro del área de I/D Procesamiento Distribuido y Paralelo y se enmarca dentro del proyecto de investigación: Evaluación de arquitecturas distribuidas de bajo costo basadas en software libre, que ha sido presentado oportunamente para el período 2015-2016 y está en proceso de evaluación. A pesar de ello se viene trabajando en proyectos afines que tuvieron como unidades ejecutoras al Departamento e Instituto de Informática de la FCEFYN de la UNSJ. El grupo de investigación viene trabajando desde hace más de 16 años con numerosas publicaciones y formación de recursos humanos en el área.

Introducción

La computación distribuida es un fenómeno de un crecimiento notable en los últimos años. La reducción de costos de hardware, los avances en tecnología de comunicación, el crecimiento explosivo de Internet y la dependencia cada vez más creciente de aplicaciones de red que van desde comunicación social hasta aplicaciones financieras, han contribuido a este crecimiento. El surgimiento de los sistemas embebidos, la nanotecnología y los sistemas inalámbricos, han abierto una nueva frontera para aplicaciones como

Internet de las Cosas, computación móvil, entre otras. La rápida adopción de Cloud Computing y la importancia creciente de big data han cambiado el panorama que presenta la computación distribuida.

Su aparición se debe a la necesidad de resolver problemas demasiado grandes para cualquier computadora, manteniendo la flexibilidad de trabajar en múltiples problemas más pequeños [1].

Los sistemas distribuidos proporcionan de forma transparente la compartición de recursos sean locales o remotos, facilitando el acceso y la gestión, e incrementando la eficiencia y la disponibilidad, a un costo más asequible.

Son sistemas en los cuales componentes conectados a través de una red de computadoras se comunican y coordinan sus acciones mediante el intercambio de mensajes[2]. Es una colección de computadoras independientes que dan la apariencia al usuario de ser una computadora única [3].

Existen diferentes tipos de sistemas distribuidos: cómputo distribuido, almacenamiento distribuido y sistemas ubicuos distribuidos. En el proyecto marco de este trabajo, se enfocaran las investigaciones en los sistemas de cómputo distribuidos, los cuales permiten realizar de manera más eficiente tareas de computación de alta prestaciones basadas en el paradigma de memoria distribuida. Ejemplos de Arquitecturas que soportan este tipo de sistemas son los cluster y el cloud computing.

El término cluster [4] se aplica a los conjuntos o conglomerados de computadoras construidos mediante la utilización de componentes de hardware comunes y que se comportan como si fuesen una única computadora. Hoy en día desempeñan un papel importante en la solución de problemas de las ciencias, las ingenierías y del comercio moderno.

La tecnología de clusters ha evolucionado en apoyo de actividades que van desde aplicaciones de super-cómputo y software de misiones críticas, servidores web y comercio electrónico, hasta bases de datos de alto rendimiento, entre otros usos.

El cómputo con clusters surge como resultado de la disponibilidad de procesadores económicos de alto rendimiento, redes de alta velocidad, el desarrollo de herramientas de software para cómputo distribuido de alto rendimiento, además es impulsado por la creciente necesidad de potencia computacional para variadas aplicaciones. Por su parte cloud computing es un modelo que permite el acceso en red omnipresente, conveniente y bajo demanda de 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.

El NIST ha presentado una de las definiciones de Cloud más aceptadas. 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” [5] 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.

Cloud está organizada en una arquitectura de capas llamada Modelo de Servicio, donde SaaS (Software as a Service) se encuentra en el tope de la pila y permite a los usuarios tener acceso a las aplicaciones alojadas en el cloud (por ej.: Google App). 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 ej.: Google App Engine). Por último, IaaS (Infraestructura as a Service) ofrece una infraestructura virtualizada de recursos (procesador, red, base de datos, disco, etc.) como servicio.

La 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 la misma, es posible compartir recursos, permitiendo la asignación y reasignación 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.

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 [6] [7].

También es importante destacar que es posible la construcción de arquitecturas híbridas que permitan montar cluster como servicio (CaaS), permitiendo un mejor aprovechamiento de los recursos virtualizados, con el objeto de utilizar ecosistemas, como Hadoop MapReduce para procesamiento distribuido de grandes volúmenes de datos.

Tomando como base estas arquitecturas es posible convertirlas en una opción de commodities usando versiones basadas en equipamiento con recursos software y hardware limitado y conectividad tradicional. De esta manera será posible evaluar el comportamiento de estas arquitecturas en función de los diferentes algoritmos que se corran sobre ellas.

Líneas de Investigación, Desarrollo e Innovación

La mayoría de las aplicaciones de computación distribuida se centran alrededor de un conjunto de sub problemas principales. El conocimiento de estos sub problemas requiere de una base de la teoría subyacente y de numerosos aspectos algorítmicos.

Los algoritmos de álgebra lineal, al igual que los algoritmos de reparto de carga, son utilizados en computación distribuida y paralela para evaluar la performance de la distribución de datos y el cómputo. La variedad de este tipo de algoritmos permite calcular la eficiencia de los mismos para distintos niveles de granularidad y tráfico de datos.

Las organizaciones de tamaño medio o pequeño (organismos descentralizados del estado, PYMES, departamentos e institutos) necesitan mantener sus datos

privados dentro de la institución. Por lo tanto el modelo híbrido (datos en el propio centro de datos y el resto en la nube), adquiere mayor aceptación.

A pesar de la popularidad de la computación distribuida, existen escasos trabajos en áreas que asocien cloud computing con Cluster, en un sentido eficiente, para propósitos de centro de datos pequeños, cloud privados y otros sistemas distribuidos de bajo costo.

La computación distribuida presenta soporte para HPC. Esta última tiene dos líneas, que son: simulación numérica y big data, y se aplicarán algoritmos para ambos líneas como parte de las evaluaciones de la presente propuesta de proyecto de investigación.

Evaluar, analizar y construir una arquitectura de bajo costo que sea eficiente en tiempo y uso de recursos, no es tarea sencilla. La complejidad de una arquitectura distribuida está dada por la diversidad de componentes que la misma incluye como son: almacenamiento, servidores, arquitecturas de software en capas, variedad de middlewares y redes, que deben trabajar conjuntamente de la forma más óptima posible y además deben ser configurados adecuadamente para ser utilizados bajo demanda.

Los diferentes Sistemas Distribuidos presentan infraestructura que está caracterizada por diferentes atributos tales como precio, performance, robustez, consumo de energía (green computing), y por lo tanto las tareas de programación y reparto de tareas se hacen más complejos y desafiantes. Esto se complica aún más cuando se trabaja sobre supuestos de distribución de tareas multicriterio.

Algunas aplicaciones son altamente escalables y pueden correr eficientemente en ambientes de Cloud estándar, mientras que otras son dependientes de una comunicación rápida entre nodos de un Cluster dedicado y posiblemente una

arquitectura de HPC en el medio. Además la latencia afecta notablemente la escalabilidad y otros parámetros.

Algunos problemas a resolver son inherentes a los sistemas distribuidos como los requerimientos de escalabilidad, fiabilidad y disponibilidad. Al menos, parte de alguno de estos sistemas deben ser locales, puesto que existe una diferencia de microsegundos para datos locales a milisegundos para datos remotos. Por otro lado, algunas tareas se benefician de la no localidad. Por ej., tener datos duplicados en otro lugar permite la continuidad del negocio cuando un data center sale de línea.

Se deben redefinir los modelos de servicios de MapReduce en Cloud para proveer servicios de bajo costo, en contraste con el servicio tal como el que ofrece un Cloud genérico de cómputo o un Cloud de MapReduce dedicado. El mismo debe manejar adecuadamente la carga de trabajo, que tiene una significativa cantidad de trabajos interactivos, pero además presenta recursos limitados.

Los sistemas distribuidos de alto costo, suelen tener recursos casi ilimitados que incluyen sistemas de archivos de altas prestaciones como Lustre, PanFS o General Parallel File Systems, redes de muy alta velocidad como InfiniBand o Ethernet a 10Gbps, soporte para HPC con placas Nvidia Tesla o similar, administradores de recursos distribuidos como Terascale Open-source Resource and QUEUE Manager, librerías específicas para simulación científica como BT Life Science HPC o Computational Fluid Dynamic. Estas alternativas no son soportadas por sistemas distribuidos de commodity ya sea por falta de recursos de hardware o costos.

Resulta por lo tanto, necesario evaluar las diferentes plataformas de Cloud y/o Cluster adecuadas para Cloud privados o

centro de datos que resuelvan eficientemente la optimización de los recursos, para sistemas distribuidos de bajo costo con limitaciones de recursos y basados en componentes de Software libre.

Resultados y Objetivos

Resultados Obtenidos

Durante los últimos cinco años se trabajó en proyectos sobre Cloud Computing y en particular durante los últimos dos años sobre Cloud híbridos. La experiencia sobre los Cloud privados, junto con líneas de investigación anteriores [8], impulsó esta línea de investigación. El grupo ha realizado nueve publicaciones en el área durante el último año: tres trabajos en el WICC 2015, un trabajo en el CACIC 2015, dos trabajos en las Jornadas de Cloud Computing, además se realizaron tres publicaciones en revistas científicas. Se han aprobado tres tesinas de grado y un trabajo de especialización.

Objetivos

El objetivo del grupo de investigación es realizar la evaluación del desempeño y otros parámetros en arquitecturas distribuidas de commodity, analizando además como impacta la virtualización y la migración de aplicaciones y servicios entre plataformas.

Formación de Recursos Humanos

El equipo de trabajo está compuesto por los seis docentes-investigadores que figuran en este trabajo y cuatro alumnos. Se están realizando cuatro tesinas de licenciatura una sobre evaluación de algoritmos de algebra lineal sobre arquitecturas diversas, otra sobre Cloud Computing Privado, otra sobre dispositivos de juegos aplicados a salud y

otra sobre SOA aplicada a Cloud. Se espera realizar también una tesis de maestría sobre Metodologías de desarrollo aplicadas a SaaS, otra sobre bases de datos NoSQL y otra sobre algoritmos de Cómputo Intensivo para Big Data y su implementación en Clouds. Además aumentar el número de publicaciones. Por otro lado también se prevé la divulgación de varios temas investigados por medio de cursos de postgrado y actualización o publicaciones de divulgación.

Referencias

- [1] M. L. Liu – Distributed Computing Principles and Applications – Pearson Addison Wesley. (2004).
- [2] George Coulouris, Jean Dollimore, Tim Kindberg y Gordon Blair - Distributed Systems Concept and Design 5a Edition - Pearson Educación. (2012).
- [3] A. Tanenbaum, M. Van Steen – Sistemas Distribuidos Principios y Paradigmas – 2ª Edición – Pearson Prentice Hall. (2008).
- [4] R. Buyya - High Performance Cluster Computing. New Jersey: F'rentice. (1999).
- [5] Mell, P., Grance, T.. "The NIST definition of cloud computing". NIST Special Publication 800 – 145 (2011).
- [6] Murazzo, M., Rodríguez, N., Medel, D., Chávez, S., Martin, A., Valenzuela, F. "Análisis de mejora en la escalabilidad de las infraestructuras de cloud computing". III JCC & big data. UNLP (2015).
- [7] Murazzo, M., Tinetti, F., Rodríguez, N. "Despliegue de una Infraestructura Cloud Privada de Código Abierto". III JCC & big data UNLP (2015).
- [8] Rodríguez N., Murazzo M., Villafaña D., Alves M., Medel D. "Integración de Computación Heterogénea con Hadoop para Cloud Computing". WICC 2013. UADER. Paraná (2013)