

Utilización Eficiente y Sostenible de Recursos en Clúster de Bajo Costo.

Karina Cenci, José Moyano, Andres Salamanca

Laboratorio de Investigación en Sistemas Distribuidos (LISiDi)
Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur
Bahía Blanca, Argentina
e-mails: {kmc, jose.moyano, cas}@cs.uns.edu.ar

Resumen

La computación en la nube permite en forma conveniente y bajo demanda, el acceso a un conjunto de recursos compartidos y personalizables que pueden ser rápidamente suministrados y liberados. La inmersión y omnipresencia de las computadoras en el mundo requiere que consideremos cómo impactan en el mismo. La sostenibilidad es uno de los aspectos primordiales al momento de considerar soluciones, ya que no sólo involucra al medioambiente sino integralmente a un conjunto de dimensiones, como la tecnología.

En la actualidad, la gran mayoría de las aplicaciones, están pensadas y desplegadas para que funcionen en entornos web y móviles. El éxito está relacionado con el tiempo de respuesta, escalabilidad y disponibilidad. Para alcanzar estas metas, es necesario contar con sistema distribuido que gestione el despliegue de los servicios.

A medida que las necesidades de recursos computacionales se trasladan de los dispositivos de los usuarios hacia los grandes *datacenters*, y que el crecimiento de estos centros de cómputo crece en proporción a su demanda, la sociedad cuestiona la sustentabilidad del modelo. El impacto al medioambiente que tienen los centros de cómputo está relacionado a la energía y el espacio requeridos, así como los desperdicios que producen los

repuestos y las actualizaciones tecnológicas.

Palabras clave: Sistemas Distribuidos, Contenedores, Eficiencia, Sostenible

Contexto

Esta línea de investigación y desarrollo se realiza en el ámbito del Departamento de Ciencias e Ingeniería de la Computación de la Universidad Nacional del Sur (UNS). En particular, como parte de las tareas que se realizan en el Laboratorio de Investigación en Sistemas Distribuidos (LISIDI). El proyecto se financia parcialmente con fondos asignados por la UNS a proyectos de investigación.

1. Introducción

La omnipresencia y ubicuidad de los dispositivos en la vida cotidiana hacen que la sociedad consuma los servicios que ofrecen los distintos organismos de agencias públicas y privadas. La computación en la nube (Cloud Computing) permite en forma conveniente y bajo demanda, acceso a la red sobre un conjunto de recursos informáticos compartidos y personalizables que pueden ser rápidamente suministrados y liberados con un mínimo esfuerzo de gestión o interacción del proveedor de servicios. Estos sistemas permiten a miles de computadoras distribuidas por Internet ser compartidas por los usuarios,

teniendo en cuenta aspectos de escalabilidad, seguridad, disponibilidad, tolerancia a fallas, sistemas operativos, apoyo a la programación paralela, servicios de búsqueda, administración de sistemas, reserva de recursos, heterogeneidad y muchos más.

La inmersión de las computadoras en el mundo requiere que consideremos el impacto que tiene en el mundo. Uno de los aspectos en los cuales es importante trabajar es en la sostenibilidad, que no sólo involucra al medioambiente sino integralmente a un conjunto de dimensiones. La sostenibilidad se la define como un concepto sistémico y debe entenderse en un conjunto de dimensiones, que incluyen social, individual, ambiental, económica y técnica. La tecnología es parte del dilema para el desarrollo de soluciones sostenibles para mejorar el estado actual y futuro de nuestro mundo.

En la actualidad, la gran mayoría de las aplicaciones, especialmente aplicaciones web y móviles, centran su funcionamiento en una arquitectura cliente-servidor, donde la aplicación web o *app* móvil es el cliente de un sistema preparado para servir desde miles hasta millones de clientes. El éxito de estas aplicaciones depende en igual proporción de satisfacer tanto requerimientos funcionales como no funcionales. Entre los requisitos no funcionales, resultan fundamentales el tiempo de respuesta, la escalabilidad y la disponibilidad. Para lograr este conjunto de requerimientos no funcionales, la aplicación se despliega sobre un sistema distribuido, construido como un clúster, manteniendo réplicas del servicio. Estas réplicas mejoran el tiempo de respuesta al balancear las solicitudes entre cada una de ellas; permiten una alta disponibilidad, al contar con múltiples copias que pueden reemplazar una en caso de fallo; y facilita la escalabilidad con la

incorporación de recursos físicos adicionales al clúster, para aumentar las réplicas y la capacidad.

A medida que las necesidades de recursos computacionales se trasladan de los dispositivos de los usuarios hacia los grandes *datacenters*, y que el crecimiento de estos centros de cómputo crece en proporción a su demanda, la sociedad cuestiona la sustentabilidad del modelo. El impacto al medio ambiente que tienen los centros de cómputo está relacionado a la energía y el espacio que requieren para funcionar, así como los desperdicios que producen los repuestos y las actualizaciones tecnológicas. Una de las posibles formas de abordar esta problemática se encuentra en mejorar la sustentabilidad de hardware. Esta mejora se propone a partir de la migración de las plataformas físicas convencionales hacia *System-on-Chip* (SoC) y *single-board computers*. Estas dos opciones mejoran significativamente el desempeño de la plataforma en relación a la cantidad de energía consumida, requieren mucho menos espacio para operar, y reducen la cantidad de desperdicios. Sin embargo, este enfoque presenta nuevos desafíos técnicos que deben ser abordados.

El resto de este artículo se organiza de la siguiente manera: la sección 2 explica los objetivos de investigación, la sección 3 los resultados esperados/obtenidos y, por último, la sección 4 discute la formación de recursos humanos.

2. Líneas de Investigación y Desarrollo

El estudio de los sistemas distribuidos y paralelos, los paradigmas de computación en la nube, *peer-to-peer*, y movilidad han adquirido gran importancia en distintos ambientes como son los sistemas orientados a la

información y comunicación, los sistemas de censado del ambiente como la temperatura y humedad, para mejorar la producción y calidad de vida. En este proyecto se van a seguir dos líneas de investigación:

- Utilización de contenedores para la extensibilidad y escalabilidad de servicios en una nube privada sobre un clúster con dispositivos de pequeña escala. El estudio, análisis y desarrollo de un *framework* que soporte el diseño y la implementación de servicios que puedan replicarse haciendo uso eficiente de los recursos de manera que balancee las distintas cargas como, por ejemplo: procesamiento, transferencia de datos, consumo energético.
- Automatización para la recuperación y tolerancia a fallas en un clúster con dispositivos de pequeña escala. El mantenimiento y actualización de clúster requiere la utilización de herramientas que faciliten el control y verificación del estado de los componentes. La estabilidad de la infraestructura es esencial para la disponibilidad de los servicios y para el uso eficiente de los mismos.

El contexto de estos desarrollos se puede aplicar en organizaciones, empresas de pequeña y mediana escala para mejorar los procesos del negocio. La utilización de contenedores e infraestructura de bajo costo facilita el testeo y pruebas en entornos similares a los de producción. Además, la facilidad de brindar servicios que estén distribuidos y replicados promueve el uso eficiente y sostenible de los recursos.

3. Resultados Obtenidos/Esperados

El diseño e implementación de la plataforma del clúster es el primer paso realizado para el desarrollo de las actividades [7]. El proceso de construcción de esta plataforma conlleva varias etapas: *inicio, en funcionamiento y automatización*. Las metas requeridas son:

- **Sustentabilidad:** Hardware de bajo costo de adquisición, costo bajo de mantenimiento, bajo consumo energético, instalable en espacios reducidos o compartidos.
- **Ambiente distribuido:** la selección de la plataforma de software que provea propiedades de un ambiente distribuido.
- **Disponibilidad y Facilidad en el Mantenimiento:** diseño de procesos repetibles, robustos y autodocumentados para la instalación y monitoreo del sistema; soporte de telemetría del sistema; características preparadas para el cambio continuo.

El primer ciclo en el diseño de la infraestructura se encuentra en el diseño de testeos para verificar la automatización y la deuda de hardware, algunos de los inconvenientes que presentan este tipo de dispositivos es la fragilidad y la limitada robustez de los mismos, que en algunos casos está asociado al bajo costo. La estabilidad y funcionamiento de la infraestructura permite avanzar en el diseño de la capa de software como middleware soporte para la alcanzar disponibilidad y tolerancia a fallas.

Los dispositivos utilizados tienen restricciones para la ejecución de aplicaciones, los primeros despliegues para el análisis de la replicación de servicios se están realizando con contenedores, específicamente con Docker swarm.

4. Formación de Recursos Humanos

En relación con la formación de los recursos humanos, el trabajo a desarrollar permite una capacitación práctica adecuada de los profesionales involucrados, permitiendo plasmar con logros concretos las investigaciones realizadas. Además, esta línea de investigación permitirá la dirección de tesis de licenciatura y trabajos finales de ingeniería afines. Asimismo, servirán para generar courseware para materias optativas a dictarse en el Departamento de Ciencias e Ingeniería de la Computación de la UNS.

5. Referencias

- [1.] Bashari Rad, Babak & Bhatti, Harrison & Ahmadi, Mohammad. (2017). An Introduction to Docker and Analysis of its Performance. *IJCSNS International Journal of Computer Science and Network Security*. 173. 8.
- [2.] W. Felter, A. Ferreira, R. Rajamony and J. Rubio, "An updated performance comparison of virtual machines and Linux containers," *2015 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*, Philadelphia, PA, USA, 2015, pp. 171-172, doi: 10.1109/ISPASS.2015.7095802.
- [3.] Sheng Li, Kevin Lim, Paolo Faraboschi, Jichuan Chang, Parthasarathy Ranganathan, and Norman P. Jouppi. 2011. System-level integrated server architectures for scale-out datacenters. In *Proceedings of the 44th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO-44)*. Association for Computing Machinery, New York, NY, USA, 260–271. <https://doi.org/10.1145/2155620.2155651>
- [4.] R. Morabito, "Power Consumption of Virtualization Technologies: An Empirical Investigation," *2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing (UCC)*, Limassol, Cyprus, 2015, pp. 522-527, doi: 10.1109/UCC.2015.93.
- [5.] Amit M Potdar, Narayan D G, Shivaraj Kengond, Mohammed Moin Mulla, Performance Evaluation of Docker Container and Virtual Machine, *Procedia Computer Science*, Volume 171, 2020, Pages 1419-1428, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2020.04.152>.
- [6.] Shuja, Junaid & Bilal, Kashif & Madani, Sajjad & Othman, Mazliza & Ranjan, R. & Balaji, Pavan & Khan, Samee. (2016). Survey of Techniques and Architectures for Designing Energy-Efficient Data Centers. *IEEE Systems Journal*. 10. 507-519. 10.1109/JSYST.2014.2315823.
- [7.] J. Moyano, C. A. Salamanca, K. Cenci, J. Ardenghi. Evaluación de Docker en un Cluster de Bajo Costo y Alta Disponibilidad, en 10mo Congreso Nacional de Ingeniería Informática y Sistemas de Información (CoNaIISI). *AJEA*, (17), 1–2021.
- [8.] R. Buyya, C. Yeo, and S. Venugopal, "Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities," in *High Performance Computing and Communications, 2008. HPCC'08. 10th IEEE International Conference on*. IEEE, 2008, pp. 5–13.
- [9.] Imtiaz Ahmad, Mohammad Gh. AlFailakawi, Asayel AlMutawa, Latifa Alsalman, Container scheduling techniques: A Survey and assessment, *Journal of King Saud University - Computer and Information Sciences*, Volume 34, Issue 7, 2022, Pages 394-3947, ISSN 1319-1578, <https://doi.org/10.1016/j.jksuci.2021.03.002>
- [10.] An Antonio Jose Alves Neto, Jose Aprigio Carneiro Neto, and Edward David Moreno Ordonez. 2022. Low-cost clusters on big data - A systematic study. In *11 Euro American Conference on Telematics and Information Systems (EATIS2022)*, June 1–3, 2022, Aveiro, Portugal. ACM, New York, NY, USA 7 Pages. <https://doi.org/10.1145/3544538.3544635>.
- [11.] Bibal Benifa J.V., Dejeay Dharma, HAS: Hybrid auto-scaler for resource scaling in cloud environment, *Journal of Parallel and Distributed Computing*, Volume 120, 2018, Pages 1-15, ISSN 0743-7315, <https://doi.org/10.1016/j.jpdc.2018.04.016>.
- [12.] Luiz F. Bittencourt, Alfredo Goldman, Edmundo R.M. Madeira, Nelson L.S. da Fonseca, Rizos Sakellariou, Scheduling in distributed systems: A cloud

- computing perspective, *Computer Science Review*, Volume 30, 2018, Pages 31-54.
- [13.] Botma, Eugene and Eduan Kotzé. "Feasibility of a low-cost computing cluster in comparison to a high-performance computing cluster: A developing country perspective." *CONF-IRM (2016)*.
- [14.] Carlos de Alfonso, Amanda Calatrava, Germán Moltó, Container-based virtual elastic clusters, *Journal of Systems and Software*, Volume 127, 2017, Pages 1-11, ISSN 0164-1212.
- [15.] Divya. V, Leena Sri. R, Docker based Intelligent Fall Detection using Edge-Fog Infrastructure, *IFAC-PapersOnLine*, Volume 53, Issue 1, 2020, Pages 45-50, ISSN 2405-8963.
- [16.] Rekha Gangula, G. Jagadeesh Kumar, Sesham Anand, A.V. Krishna Prasad, Integration of dynamic Docker containers and kubernetes with advanced cloud and Internet of Things, *Materials Today: Proceedings*, 2021, ISSN 2214-7853, <https://doi.org/10.1016/j.matpr.2021.07.276>
- [17.] Monika Gill, Dinesh Singh, ACO Based Container Placement for CaaS in Fog Computing, *Procedia Computer Science*, Volume 167, 2020, Pages 760-768, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2020.03.406>.
- [18.] Dominik Göddeke, Dimitri Komatitsch, Markus Geveler, Dirk Ribbrock, Nikola Rajovic, Nikola Puzovic, Alex Ramirez, Energy efficiency vs. performance of the numerical solution of PDEs: An application study on a low-power ARM-based cluster. *Journal of Computational Physics*, Volume 237, 2013, Pages 132-150, ISSN 0021-9991, <https://doi.org/10.1016/j.jcp.2012.11.031>.
- [19.] H. Jiang, K. Wang, Y. Wang, M. Gao and Y. Zhang, "Energy big data: A survey," in *IEEE Access*, vol. 4, pp. 3844-3861, 2016, doi: 10.1109/ACCESS.2016.2580581.
- [20.] Kumari, Priti and Parmeet Kaur. "A survey of fault tolerance in cloud computing." *J. King Saud Univ. Comput. Inf. Sci.* 33 (2021): 1159-1176.
- [21.] A. Martin, S. Raponi, T. Combe, R. Di Pietro, Docker ecosystem – Vulnerability Analysis, *Computer Communications*, Volume 122, 2018, Pages 30-43, ISSN 0140-3664, <https://doi.org/10.1016/j.comcom.2018.03.011>.
- [22.] Partha Pratim Pande and Amlan Ganguly. 2012. Introduction to the special issue on sustainable and green computing systems. *J. Emerg. Technol. Comput. Syst.* 8, 4, Article 26 (October 2012), 3 pages. <https://doi.org/10.1145/2367736.2367737>