

Tecnologías HPC Híbridas en la Solución de Problemas Complejos

Mercedes Barrionuevo ⁽¹⁾, Francisco Constantini⁽²⁾, Julian Escalante ⁽²⁾, Mariela Lopresti ⁽¹⁾,

Maximiliano Lucero ⁽¹⁾, Natalia Miranda ⁽¹⁾ y Fabiana Piccoli ^(1 y 2).

⁽¹⁾LIDIC- Univ. Nacional de San Luis, San Luis

⁽²⁾Universidad Autónoma de Entre Ríos, Concepción del Uruguay

⁽³⁾ Univ. Nacional de San Juan, San Juan

Argentina

{mbarrio, omlopres, mlucero, ncmiran, mpiccoli}@unsl.edu.ar

{franciscostantini,escalantejulian18}@gmail.com

RESUMEN

Los problemas a resolver hoy en día mediante soluciones computacionales demandan muchos recursos, esto puede obedecer a dos factores: el tamaño del problema generado por los datos que involucra, o su complejidad. Cualquiera sea el caso, siempre los resultados se esperan en un tiempo razonable. Una de las soluciones propuestas es pensar en sistemas heterogéneos: una computadora formada por procesadores many y multicores con software capaz de tomar ventaja de cada una de las componentes subyacentes.

En este trabajo se exponen dos líneas de trabajo orientadas a distintos tipos de problemas, en las cuales se propone desarrollar técnicas de Computación de Alto Desempeño para el escenario descrito.

Palabras clave: Computación de Alto Desempeño. Modelos Híbridos. Problemas Complejos. Big Data.

CONTEXTO

Esta propuesta de trabajo se lleva a cabo dentro de los proyectos de investigación: “Tecnologías Avanzadas aplicadas al Procesamiento de Datos Masivos” (LIDIC, UNSL) y “Cómputo de Altas Prestaciones

aplicado a la Solución de Grandes Problemas” (UADER).

1. INTRODUCCIÓN

Los datos, las soluciones complejas y los problemas a resolver computacionalmente han crecido exponencialmente en la última década. Por esta razón tanto el hardware como el software tuvieron que evolucionar ajustándose a los nuevos requerimientos.

Los sistemas de computación híbridos permiten combinar las características de las arquitecturas multi-core [Bas16] y many-core [Aam18, Hon10], los sistemas de memoria compartida [Vaj11] y de memoria distribuida [Kau14, Mar17], incrementan la capacidad y poder de cómputo del hardware del sistema de computación resultante. Esto posibilita aprovechar las ventajas ofrecidas: múltiples procesos y threads con distintas administraciones de memoria coexistiendo en el mismo sistema.

Si se considera el software, se han desarrollado diferentes paradigmas teniendo en cuenta las características de estos nuevos sistemas. Si bien los primeros desarrollos se enfocaron en los paradigmas típicos de paralelismo de datos y de tareas [Pac11]; o según la arquitectura: modelo de memoria compartida (OpenMP[Op19]), de memoria

distribuida o pasaje de mensajes (MPI[Gro14]) y para placa GPU (CUDA[NVi19]); hoy ya existen propuestas donde nos permiten combinar todas estas herramientas o independizarnos de los dispositivos utilizado como co-procesador alguna de las placas: GPU Intel/Radeon/NVidia, FPGA[Sk119], arduino, etc. a través de una única herramienta. Esto es posible utilizando tecnologías de alto nivel para desarrollar aplicaciones multiplataforma como son openCL (Open Computing Language)[Kae15] y oneAPI[One20]. Ambas tecnologías permiten desarrollar soluciones computacionales portables independientes del proveedor y del dispositivo, capaces de lograr tiempos de respuestas menores en muchas plataformas de hardware diferentes. Ambas herramientas permiten la programación de aplicaciones paralelas sobre unidades de computación heterogéneas, como GPUs, CPUs y otros dispositivos de aceleración. Su estructura consta de dos partes: Una API para acceder y usar las unidades de computación, y un lenguaje de programación que se ejecutará dentro del dispositivo o unidad de proceso. [Kae15, One20].

Teniendo en cuenta la demanda de cómputo de los problemas actuales y la posibilidad que nos brindan tanto las arquitecturas híbridas como los modelos de programación de alto nivel se puede lograr aplicar una estrategia de paralelización más efectiva mediante múltiples niveles de paralelismo y posibilitar así la reducción del overhead de comunicación. Existen numerosas soluciones a problemas complejos de gran escala, las cuales aplican técnicas HPC multi-tecnología y multiplataforma [Car18, Isu19].

Por todo lo expuesto, nuestra motivación es investigar, verificar y poner en marcha nuevas técnicas y arquitecturas híbridas para mejorar el procesamiento y los tiempos de respuesta de diferentes aplicaciones complejas. En base a ello, nuestro objetivo es aplicar HPC

multi-tecnología para resolver problemas complejos en ambientes híbridos. En la próxima sección presentamos los diferentes casos de estudio considerados

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

Varios son los sistemas complejos considerados como benchmark para aplicar técnicas de HPC híbridas. Se destacan:

- Problemas de Simulación de Sistemas Reales: La toma de decisiones en sistemas compuestos por múltiples componentes interrelacionadas, no es una tarea fácil. Una posibilidad es abordarlos analíticamente o mediante ensayos o experimentos, lo cual puede implicar mucho riesgo. Una buena solución suele ser analizarlos a través de técnicas de simulación por computadora, las cuales representan una de las herramientas más poderosas para la resolución de problemas. Su potencial es incalculable, ya que permiten abstraer sistemas del mundo real y tomar decisiones basadas en múltiples experimentos. Los Autómatas Celulares (AC) son modelos matemáticos muy útiles y adecuados, los cuales a través de un conjunto de reglas simples pueden representar sistemas complejos pertenecientes a diferentes áreas científicas, como química, bioquímica, economía, física, etc.[Kau84, Wol84]. Además tienen como característica, su naturaleza paralela. En función de todas estas características, hemos enfocado el estudio de varios sistemas a través de AC paralelos y en distintas tecnologías. Ellos son:
 - El contagio de enfermedades como la Gripe A y Covid-19. Para ambos casos se desarrolló un AC donde se simuló un ambiente social dinámico y las reglas para cada una de las

- enfermedades[Hup13, Joh09, Wor14]. La propuesta fue desarrollada para arquitecturas many-core, tanto Nvidia, AMD e Intel, utilizando las herramientas de programación provistas por CUDA y OpenCL. Los desarrollos nos permiten estudiar, en menos tiempo, el proceso de propagación teniendo en cuenta diferentes realidades: el tipo de población, su distribución y otras características; a fin de tomar decisiones, como las campañas de vacunación, el aislamiento, cuarentena, mecanismo de contención, etc.. Los resultados fueron validados con los reportados por la OMS.
- Difusión de Noticias/Rumores: Este trabajo plantea, mediante un simulador con AC y técnica HPC, analizar el comportamiento de una sociedad frente a la dispersión de una noticia o rumor [Mai13, Yir12]. También se definió el AC, especificando los estados y reglas de transición. En la implementación se aplicaron técnicas de HPC para memoria distribuida. Esto nos habilita a realizar un análisis del comportamiento de una sociedad, frente a una noticia o rumor, siendo de utilidad tanto para fines buenos como programas de concientización o campañas de solidaridad, de política, etc., o no buenos como las fake news, campañas políticas, de desprestigio, etc.. De los resultados preliminares se pudo observar cómo estas últimas tienen mayor difusión.
- La versión actual está siendo ajustada y extendida a fin de incorporar, no sólo otras características relacionadas al problema cómo la difusión virtual, sino también la inclusión de otras tecnologías HPC.

- Otros problemas: En este caso, se están analizando distintos problemas a fin de simular su comportamiento, para ello estamos en la etapa de investigación de la dinámica del comportamiento de los siguientes fenómenos: Desarrollo de la roya en el trigo (Agricultura); Desplazamiento de los bancos arenas en el río Uruguay (Hidrología) y Crecimiento demográfico de una región para su buena planificación urbana (Sociedad). La solución presentada tendrá la característica de aplicar técnicas HPC híbridas utilizando las nuevas herramientas como oneAPI.
 - Problemas con Grandes Volúmenes de Datos: Muchas aplicaciones de Internet generan un tráfico de datos con características específicas susceptibles de clasificación utilizando Aprendizaje de Máquina. Las técnicas de aprendizaje automático guiados sin supervisión son un camino prometedor para hacer frente a los constantes cambios en el tráfico de red.
- El principal obstáculo al que nos enfrentamos al momento de avanzar en la clasificación del tráfico es un problema persistente de la investigación de Internet en general: la falta de una variedad de huellas que se puedan compartir para servir como datos de prueba, así como la ausencia de objetos de flujo utilizados como referencia para su validación. Estamos desarrollando un sistema paralelo de detección de ataques y anomalías circulantes en una red de computadoras, buscando llevar a cabo la tarea de identificación de ataques a nivel de red y transporte del modelo TCP/IP.
- La tarea de detectar posibles paquetes de datos anómalos en una red de computadoras es muy costosa, combinar

técnicas de clasificación y visualización de datos implementadas con técnicas HPC en su solución parece una buena alternativa. Para ello se busca identificar patrones que se desvían del comportamiento normal a partir del tráfico circulante en la red (gran volumen de datos) en un tiempo cercano a su ocurrencia, a fin de poder tomar decisiones rápidas [Bar16, Bar17, Bar18]. El trabajo está siendo desarrollado para GPU Nvidia.

Todas estas líneas de investigación tienen en cuenta la escalabilidad del problema y la portabilidad de los desarrollos, no sólo respecto a las arquitecturas sino también al tipo de problema.

3. RESULTADOS OBTENIDOS/ESPERADOS

Como objetivos de las líneas de investigación nos planteamos facilitar el desarrollo de soluciones paralelas portables, de costo predecible, capaces de explotar las ventajas de modernos ambientes HPC a través de herramientas y “frameworks de computación” de alto nivel. Los resultados obtenidos hasta el momento son muy satisfactorios.

4. FORMACIÓN DE RECURSOS HUMANOS

Los resultados esperados respecto a la formación de recursos humanos son hasta el momento el desarrollo de 3 tesis de posgrado: 1 de doctor y 2 de maestría; y varias tesinas de grado en las distintas universidades intervinientes.

5. BIBLIOGRAFÍA

[Aam18] Aamodt, T., Lun Fung, W. and Rogers, T. - General-Purpose Graphics

Processor Architectures Synthesis Lectures on Computer Architecture. Editor: Morgan & Claypool Publishers. ISBN: 1627056181, 9781627056182. 2018.

[Bas16] Basu, S. “*Parallel and Distributed Computing Architectures and Algorithms*”. ISBN: 8120352122, 9788120352124. PHI Learning Pvt. Ltd., 2016.

[Bar16] Barrionuevo, M. Lopresti, M., Miranda, N. Piccoli, F.. “*Un enfoque para la detección de anomalías en el tráfico de red usando imágenes y técnicas de Computación de Alto Desempeño*”. XXII Congreso Argentino De Ciencias de la Computación. CACIC 2016. Pp. 1166-1175. Octubre 2016, San Luis, Argentina.

[Bar17] Barrionuevo, M. Lopresti, M. Miranda, N. Piccoli, F. “*Un Modelo de Detección de Anomalías en una LAN usando K-NN y Técnicas de Computación de Alto Desempeño*”. Communications in Computer and Information Science” (CCIS). ISBN: 978-3-319-75213-6. Editorial: Springer. Edición: Extranjera (inglés). Año: 2017. Volumen 790. Pp 219-228.

[Bar18] Barrionuevo, M., Lopresti, M., Miranda, N. and Piccoli, F. . “*Secure Computer Network: Strategies and Challenges in Big Data Era*”. Journal of Computer Science & Technologies. Editorial Red de Universidades Nacionales con Carreras de Informática (RedUNCI) Iberoamerican Science & Technology Education Consortium (ISTEC). Vol 3, pp 248-257. ISSN: 1666-6038 (electrónico), 1666-6046 (impreso). Diciembre 2018.

[Car18] CARLA 2018, *Selected Papers High of Performance Computing 5th Latin American Conference*. Communications in Computer and Information Science, Vol. 979. ISBN: 3030162052, 9783030162054. Colombia, 2018. Springer, 2019.

[Gro14] Gropp, W., Hoefler, T., Thakur, R. and Lusk, E. “*Using Advanced MPI: Modern Features of the Message-Passing Interface*”. Computer science & intelligent systems Scientific and Engineering Computation. ISBN: 0262527634 - 9780262527637. MIT Press, 2014.

- [Hon10] Hong, S., Kim, H., Hong, S. and Kim, H.. “*An integrated GPU power and performance model*” in Proceedings of the 37th annual international symposium on Computer architecture - ISCA '10, Vol. 38, N° 3. Pp. 280-289. 2010.
- [Hup13] Huppert, A. and Katriel, G.. “*Mathematical modelling and prediction*” in infectious disease epidemiology. Clinical Microbiology and Infection. Vol 19 N°11. Pp 999-1005, 2013.
- [Isu19] ISUM 2019, *Selected Papers of Supercomputing: 10th International Conference on Supercomputing*. Communications in Computer and Information Science, Vol. 1151. ISBN: 3030380432, 9783030380434. Springer Nature. México 2019.
- [Joh09] Johnson, T. and McQuarrie, B.. “*Mathematical modeling of diseases: Susceptible-infected-recovered (sir) model*”. In University of Minnesota, Morris, Math 4901 Senior Seminar, 2009.
- [Kae15] Kaeli, D. Mistry, P. Schaa, D. Zhang, D. P “*Heterogeneous Computing with OpenCL 2.0*” Third Edition, Morgan Kaufmann, ISBN: 978-0-12-801414-1, 2015
- [Kau14] Kaur, K. and Rai, A.. “*A Comparative Analysis: Grid, Cluster and Cloud Computing*” Int. J. Adv. Res. Comput. Commun. Eng., vol. 3, no. 3, pp. 2278–1021, 2014.
- [Kau84] Kaufman, S.. “*Emergent properties in random complex automata*”. Physica D: Nonlinear Phenomena. Vol 10 N°1 Pp 145-156, 1984.
- [Mai13] Maitanmi, O., Adekunle, Y. and Agbaje, M.. “*Model-Based Cellular Automata on Spread of Rumours*”. Computer Science, Babcock University. Nigeria. 2013.
- [Mar17] Marinescu, D.. “*Cloud Computing: Theory and Practice*”. 2nd Ed. ISBN: 0128128119, 9780128128114. Morgan Kaufmann, 2017.
- [NVi19] Nvidia. “*CUDA C++ Programming Guide, Design Guide*”. https://docs.nvidia.com/cuda/pdf/CUDA_C_Programming_Guide.pdf. 2019.
- [One20] OneAPI Specification, release 1.0-rev 3. <https://spec.oneapi.com/versions/1.0-rev-3/>. 2020
- [Op19] OpenMP Architecture Review Board. “*OpenMP Application Programming Interface Specification Version 5.0*”. ISBN: 1795759887 - 978-1795759885. 2019.
- [Pac11] Pacheco, P. “*An Introduction to Parallel Programming*”, 1st ed., San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011.
- [Sk119] Skliarova, I. and Sklyarov, V. - FPGA-BASED Hardware Accelerators. Lecture Notes in Electrical Engineering, Vol. 566 ISBN 3030207218, 9783030207212. Springer, 2019
- [Vaj11] Vajda, A. “*Multi-core and Many-core Processor Architectures. In: Programming Many-Core Chips*”. Springer, Boston, MA. 2011.
- [Wol84] Wolfram, S.. “*Universality and complexity in cellular automata*”. Physica D: Non-linear Phenomena, Vol 10 N° 1. Pp 1-35, 1984.
- [Wor14] World Health Organization. *Influenza (seasonal)*. Fact sheet N211. 2014.
- [Yir12] Yiran, G. and Jinzhu, D. “*Research on Rumors Spread Based on Cellular Automata*”, in Proceedings of the 2nd International Conference on Green Communications and Networks. Vol 2. ISBN 978-3-642-35418-2. Pp 235-244 Londres, Inglaterra. Springer Heidelberg. 2012.