

Métricas de rendimiento, aplicaciones de datos masivos e inteligencia artificial en cómputo paralelo y distribuido

Javier Balladini¹, Marina Morán¹, Claudio Zanellato¹, Claudia Rozas¹, Rodrigo Cañibano¹,
Cristina Orlandi², Armando De Giusti³, Remo Suppi⁴, Dolores Rexachs⁴, Emilio Luque⁴,
Emmanuel Frati⁵

¹ Facultad de Informática, Universidad Nacional del Comahue
{javier.balladini, marina, claudio.zanellato, claudia.rozas, rcanibano}@fi.uncoma.edu.ar

² Hospital Francisco Lopez Lima - orlandi.mariacristina@gmail.com

³ Instituto de Investigación en Informática LIDI, Universidad Nacional de La Plata - degiusti@lidi.info.unlp.edu.ar

⁴ Departamento de Arquitectura de Computadores y Sistemas Operativos, Universidad Autónoma de Barcelona
{remo.suppi, dolores.rexachs, emilio.luque}@uab.es

⁵ Universidad Nacional de Chilecito, fefrati@undec.edu.ar

Resumen

Los grandes avances tecnológicos de los sistemas de cómputo paralelo y distribuido hacen viable nuevas soluciones a problemas. Por un lado, nos enfocamos en métricas de rendimiento. En particular, nos abocamos a métricas energéticas que cobraron enorme relevancia debido al gran número de unidades de procesamiento que componen los sistemas de cómputo. Por otro lado, buscamos aplicar técnicas de cómputo paralelo y distribuido para brindar soluciones en el sector salud. En especial, nos orientamos a sistemas de alertas tempranas de gravedad, basadas en inteligencia artificial. Una aplicación está destinada a Unidades de Cuidados Intensivos, que debe tratar con datos masivos, y otra aplicación tiene como fin la clasificación del nivel de gravedad de pacientes con COVID-19, que presenta una arquitectura distribuida, tolerante a fallos y de simple administración. Los trabajos se desarrollan en colaboración con otras universidades, y un hospital público de Argentina. La formación de recursos humanos en estas líneas está orientada al nivel de tecnicatura, grado, maestría y doctoral.

Palabras claves: computación de altas prestaciones, eficiencia energética, big data, salud, inteligencia artificial.

1. Contexto

Las líneas de investigación aquí presentadas están enmarcadas dentro del proyecto de investigación 04/F017 "Cómputo paralelo y distribuido: métricas de rendimiento, aplicaciones de big data e inteligencia artificial", financiado por la Universidad Nacional del Comahue (UNComa), con inicio el 01/01/2021 y finalización el 31/12/2024, acreditado por el Ministerio de Educación de Argentina.

La mayoría de los temas de investigación surgieron en proyectos anteriores acreditados. El eje de aplicaciones para la salud, que involucra temáticas de inteligencia artificial y Big Data, se desarrolla en colaboración con el Hospital Francisco López Lima de General Roca (Río Negro) y con interés del Ministerio de Salud de la Provincia de Río Negro. El eje de métricas de rendimiento de los sistemas de cómputo paralelo, se desarrolla en colaboración con el Instituto de Investigación en Informática LIDI de la Universidad Nacional de La Plata, el grupo de investigación "High Performance Computing for Efficient Applications and Simulation" de la Universidad Autónoma de Barcelona, España, y con la Universidad Nacional de Chilecito, La Rioja.

2. Introducción

Numerosas aplicaciones informáticas que requieren el procesamiento de grandes volúmenes de datos están surgiendo gracias al uso de tecnologías de cómputo paralelo y distribuido, que hasta hace unos años eran excepcionales. La masificación tecnológica y consecuente reducción de costos hizo que se pueda pensar en el desarrollo de nuevas aplicaciones que tengan un fuerte impacto social. En particular, nos interesamos principalmente en el sector de salud.

A su vez, nos enfocamos en las métricas de eficiencia computacional y energética de los sistemas de cómputo paralelo y distribuido. Desde los inicios de la computación las métricas de interés se relacionaban únicamente con la eficiencia computacional. Actualmente, las métricas de rendimiento energéticas en muchos casos superan en relevancia a la velocidad de cómputo. En particular, nos hemos orientado a la gestión del consumo energético de estos sistemas, para reducir su impacto económico, medioambiental y social.

A continuación se introduce la motivación y problemática de ambas líneas de investigación.

2.1. Aplicaciones para la salud

Unidad de Cuidados Intensivos

Una Unidad de Cuidados Intensivos (UCI) provee cuidados continuos y rigurosos a personas adultas críticamente enfermas que pueden beneficiarse de tratamiento, y da un buen morir a pacientes irrecuperables. Los datos de los pacientes involucran datos clínicos de baja frecuencia y flujos de datos fisiológicos de alta frecuencia generados por el equipamiento médico. En una UCI típica, los enfermeros completan manualmente datos en formularios, registrando datos clínicos y fisiológicos. Los datos fisiológicos se obtienen por observación de las pantallas del equipamiento médico a intervalos de tiempo que definen los médicos para cada paciente. El equipamiento médico emite alertas cuando hay riesgo en la salud del paciente basándose en mediciones de parámetros. Luego, los médicos analizan los datos de los formularios y dan

a los enfermeros indicaciones de tratamientos.

Los principales problemas que ocurren en las UCIs típicas son: (a) gestión de la información es proclive a errores humanos, (b) pérdida de datos entre registros de enfermería, (c) detección tardía del deterioro de la salud de los pacientes. Las causas tienen origen en el tratamiento manual de la información (lento y propenso a errores), y cuyos efectos incluyen diagnósticos imprecisos o incorrectos o retrasados, inconsistencias en la información, menos datos para investigaciones médicas, y mayores necesidades de recursos humanos. En la mayoría de las UCIs del país, no hay médicos intensivistas durante las 24hs, y hay sobrecarga laboral, impactando negativamente en la salud de los pacientes.

Las soluciones se orientan a sistema expertos de detección automática, temprana y progresiva del deterioro de pacientes, tal como los propuestos en [1, 10, 13].

Detección de riesgo en pacientes con COVID-19

El 11 de marzo de 2020 la Organización Mundial de la Salud dictaminó la pandemia de COVID-19. El sistema sanitario argentino tiene su límite crítico en la cantidad de enfermeros disponibles por habitantes, con un número de 2,6 enfermos cada mil habitantes en contra parte de los 5,7 que tienen España e Italia¹, quienes no han podido contener la situación con ese personal. Esta situación generó que los recursos humanos especializados escaseen.

El “triage” es un método que permite la organización y optimización de los recursos en situaciones críticas, seleccionando y clasificando a los pacientes en diferentes niveles de gravedad para una correcta asignación de la atención. Este método fue aplicado en China [11] en pacientes con COVID-19 por medio de la adaptación de uno de los sistemas de alertas tempranas más conocido y validado. Esto permite mejorar el rendimiento del personal, reduciendo los controles en pacientes estables y

¹Datos relevados en 2017 por el Banco Mundial (<https://datos.bancomundial.org>).

aumentando los mismos en pacientes graves, y reducir la mortalidad hospitalaria inesperada en áreas de internación general.

Si bien el triage puede ser realizado mediante cálculos hechos a mano, no solo se agregaría una nueva tarea al personal de salud sino que sería un método muy propenso a errores humanos y por ende poco fiable. En consecuencia, estamos desarrollando un sistema informático inteligente que implemente un sistema de alerta temprana para pacientes en internación general con COVID-19, que realiza un triage a través del análisis automático y en tiempo real de datos clínicos, clasificando a los pacientes según el riesgo, emitiendo alertas al personal de salud, y permitiendo organizar y optimizar recursos humanos. El sistema, a partir de la carga de datos de enfermería, comorbilidades, resultados radiológicos y de laboratorio, predice la gravedad de los pacientes (bajo, moderado, alto, crítico). No conocemos actualmente sistemas de este tipo.

2.2. Métricas de rendimiento - eficiencia energética en HPC

La computación de alto rendimiento (HPC) sigue aumentando su rendimiento computacional y su eficiencia energética. Por ejemplo, el superordenador Fugaku, la supercomputadora más rápida del mundo en el ranking Top500, presenta 442 PFlops frente a los 148,6 PFlops de su predecesor, Summit. Estas máquinas, al momento de ingresar al Top500 se encontraban entre las diez de mayor eficiencia energética del ranking Green500, con alrededor de 15 GFlops/W. Sin embargo, como las máquinas aumentan enormemente de tamaño, el consumo de energía sigue incrementándose; mientras que la computadora Summit tiene un consumo de 10 MW, Fugaku tiene casi 30 MW (aproximadamente lo mismo que consume una ciudad de 400.000 habitantes). Como este aumento del consumo de energía no es sostenible, es necesario reducirlo.

La computación ecológica es el estudio y la práctica de la computación ambientalmente sostenible. Ella se ocupa de diferentes aspectos

de los sistemas de cómputo: diseño, manufactura, eliminación, y uso. Este último aspecto, el uso ecológico, se refiere al uso de los sistemas de cómputo con conciencia ambiental. Es posible reducir el consumo de energía de los sistemas de cómputo utilizando diferentes estrategias que deben ser consideradas a nivel del software [9, 6], y consisten en realizar cambios en la configuración del sistema (escalado dinámico de frecuencia y tensión, e hibernación de recursos) o en las aplicaciones (uso adecuado de la jerarquía de memoria, rediseño de algoritmos, planificación de tareas, y asignación de tareas a recursos hardware).

El aumento significativo de la cantidad de unidades de procesamiento causa el incremento del consumo de energía y la disminución de la confiabilidad del sistema de cómputo. Así, a poco tiempo de arribar a la era exaescala (o ya en ella para aplicaciones que requieren ejecutar operaciones de baja complejidad), la tolerancia a fallos y el consumo energético se han identificado como los dos mayores desafíos a enfrentar [12, 8]. En consecuencia, proponemos desarrollar metodologías, modelos y software para gestionar el consumo energético, en especial al utilizar mecanismos de tolerancia a fallos en máquinas paralelas de tipo cluster.

3. Líneas de investigación

El eje central de nuestra investigación es desarrollar metodologías, modelos y soluciones informáticas para colaborar en la resolución de problemas de cómputo paralelo y distribuido, que puedan tener una alta demanda computacional e impacto social en el campo de la salud, y en la reducción del consumo energético de sistemas de HPC.

3.1. Aplicaciones para la salud

UCI: Nuestro sistema intenta emular el comportamiento de un médico intensivista experto, dando recomendaciones para la toma de decisiones clínicas, con el objetivo de reducir la incertidumbre sobre el diagnóstico, las opciones de tratamiento y el pronóstico. La solución

requiere la aplicación de técnicas de cómputo paralelo y distribuido para procesamiento en tiempo real de algoritmos de inteligencia artificial sobre grandes volúmenes de datos. A diferencia de otras alternativas, nuestro objetivo se orienta a la construcción de un sistema multihospitalario (con el fin de incrementar el volumen de datos y consecuente extracción de conocimiento) con soporte de telemedicina, que integre componentes de software libre maduros, que nosotros optimizados para nuestro dominio, que sea seguro, tolerante a fallos y resiliente.

COVID-19: Nuestro objetivo es construir un sistema y modelo computacional de alerta temprana para pacientes con COVID-19, inicialmente basado en el conocimiento de personal experto en salud, y posteriormente mejorado mediante técnicas de aprendizaje automático. La complejidad de la arquitectura del sistema no está en la alta demanda computacional sino en evitar el uso de servidores estándares en los hospitales (a fin de simplificar la administración multihospitalaria), y que el sistema local de un hospital se mantenga operativo ante fallos de conexión con la nube, utilizando únicamente dispositivos móviles.

3.2. Métricas de rendimiento - consumo energético en HPC

Nos centramos en el desarrollo de metodologías, modelos y construcción de software para administrar y gestionar el consumo de energía y prestaciones computacionales de sistemas de cómputo paralelo. Actualmente, nuestro principal objetivo es la gestión energética en mecanismos de tolerancia a fallos basados en checkpoints.

4. Resultados y objetivos

4.1. Aplicaciones para la salud

UCI: En [5] presentamos un análisis del estado general de la UCI del hospital Francisco López Lima, y la propuesta del diseño de alto nivel del sistema. En [2] presentamos el diseño

de la infraestructura del sistema de procesamiento de reglas clínicas y un prototipo. En [3] propusimos la herramienta como mejora de la prestación de servicios integrados de cuidados intensivos de la salud. En [7] presentamos una optimización de la administración de datos de curvas fisiológicas. Próximos objetivos: continuar con el desarrollo de hardware y software para extraer datos del equipamiento médico, desarrollar aplicaciones para el procesamiento eficiente de señales, manejo de errores en datos fisiológicos, optimización de la infraestructura de datos masivos de tiempo real para el procesamiento de reglas clínicas, y el desarrollo de una aplicación para la interacción del sistema con médicos y enfermeros.

COVID-19: Estamos completando la primera versión del sistema informático (COVINDEX), que implementa modelos de alertas tempranas. El sistema se compone de una aplicación web, una aplicación móvil y una aplicación que ejecuta en un servidor en la nube. A partir de los datos históricos que se recolecten de pacientes, se calibrará periódicamente el sistema de alerta temprana de insuficiencia respiratoria, sobre la base de técnicas de aprendizaje automático, para aumentar la capacidad predictiva del sistema. Aún no se cuenta con publicaciones científicas. Próximos objetivos: estamos trabajando en conjunto con el Ministerio de Salud de la Provincia de Río Negro para construir un sistema adaptable a otras enfermedades y para el uso en pacientes con enfermedades no contempladas de manera específica, que permita derivaciones de pacientes entre hospitales y un control general de la situación por parte del ministerio.

4.2. Métricas de rendimiento - consumo energético en HPC

Los métodos de tolerancia a fallos tienen fuerte incidencia en el consumo energético de los sistemas de HPC, y resulta de suma importancia conocer, antes de ejecutar una cierta aplicación, el impacto que pueden producir los diferentes métodos y configuraciones del mismo. En [4], presentamos una metodología para

predecir el consumo energético producido por el método de checkpoint coordinado remoto, y en [15] expusimos un análisis de los factores que afectan el consumo energético de operaciones de checkpoint y restart en clusters. En [14] propusimos un modelo para estimar el consumo energético de operaciones de checkpoint y restart, y un método para su construcción. En [16] presentamos estrategias para checkpoints no coordinados que, al momento de un fallo de un nodo, permiten gestionar y reducir el consumo energético de los nodos que no han fallado; se construyó también un modelo energético y un simulador que permite evaluar las estrategias. Próximos objetivos: se está ampliando el simulador para contemplar más situaciones como las operaciones no bloqueantes y las esperas en cascada causadas por las dependencias entre procesos.

Se dictó un curso de posgrado “Análisis de rendimiento de aplicaciones paralelas” (colaboración UNCo / UNdeC) y una capacitación sobre “Eficiencia Energética en Sistemas de Cómputo” (colaboración UNCo / UNdeC / Red UNCI).

5. Formación de recursos humanos

El equipo de trabajo de la Universidad Nacional del Comahue tiene un Doctor y un Máster, una estudiante de Doctorado y dos estudiantes de Maestría. Tres estudiantes de grado están realizando trabajos de tesis, y hay un estudiante de tecnicatura. En 2021 no se ha finalizado ninguna tesis.

Referencias

- [1] S. Balaji, M. Patil, and C. McGregor. A cloud based big data based online health analytics for rural nicus and picus in india: Opportunities and challenges. In *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 385–390, 2017.
- [2] Javier Ballardini, Pablo Bruno, Rafael Zurita, and Cristina Orlandi. An automatic and early detection of the deterioration of patients in intensive and intermediate care units. *Journal of Computer Science and Technology*, 18(03):e25, Dec. 2018.
- [3] Javier Ballardini, Pablo Bruno, Rafael Zurita, Cristina Orlandi, Remo Suppi, Dolores Rexachs, and Emilio Luque. A tool for improving the delivery of integrated intensive health care performance. *International Journal of Integrated Care*, 19(4), 07/2019 2019.
- [4] Javier Ballardini, Marina Morán, Dolores Rexachs, and Emilio Luque. Metodología para predecir el consumo energético de checkpoints en sistemas de hpc. *XX Congreso Argentino de Ciencias de la Computación (CACIC 2014)*, 2014.
- [5] Javier Ballardini, Claudia Rozas, Emmanuel Frati, Nestor Vicente, and Cristina Orlandi. Big data analytics in intensive care units: challenges and applicability in an argentinian hospital. *Computer Science and Technology (JCST)*, 2015.
- [6] Aurelien Bouteiller, Franck Cappello, Jack Dongarra, Amine Guermouche, Thomas Hérault, and Yves Robert. Multi-criteria checkpointing strategies: Response-time versus resource utilization. In *European Conference on Parallel Processing*, pages 420–431. Springer, 2013.
- [7] Rodrigo Cañibano, Claudia Rozas, Cristina Orlandi, and Javier Ballardini. Data management optimization in a real-time big data analysis system for intensive care. In *Conference on Cloud Computing, Big Data & Emerging Topics*, pages 93–107. Springer, Cham, 2020.
- [8] Franck Cappello, Al Geist, William Gropp, Sanjay Kale, Bill Kramer, and Marc Snir. Toward exascale resilience: 2014 update. *Supercomputing Frontiers and Innovations*, 1(1), 2014.
- [9] Mohammed El Mehdi Diouri, Olivier Glück, Laurent Lefèvre, and Franck Cappello. Ecofit: A framework to estimate energy consumption of fault tolerance protocols for hpc applications. In *Proceedings of the 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing*, pages 522–529, 2013.
- [10] Jagreet Kaur and Dr. Kulwinder Singh Mann. AI based HealthCare platform for real time, predictive and prescriptive analytics using reactive programming. *Journal of Physics: Conference Series*, 933:012010, jan 2018.
- [11] Xuelian Liao, Bo Wang, and Yan Kang. Novel coronavirus infection during the 2019–2020 epidemic: preparing intensive care units-the experience in sichuan province, china. *Intensive care medicine*, 46(2):357–360, 2020.
- [12] Yongpeng Liu and Hong Zhu. A survey of the research on power management techniques for high-performance systems. *Software: Practice and Experience*, 40(11):943–964, 2010.
- [13] Fernando López-Martínez, Edward Rolando Núñez-Valdez, Vicente García-Díaz, and Zoran Bursac. A case study for a big data and machine learning platform to improve medical decision support in population health management. *Algorithms*, 13(4), 2020.
- [14] M. Morán, J. Ballardini, D. Rexachs, and E. Luque. Prediction of energy consumption by checkpoint/restart in hpc. *IEEE Access*, 7:71791–71803, 2019.
- [15] Marina Morán, Javier Ballardini, Dolores Rexachs, and Emilio Luque. Factores que afectan el consumo energético de operaciones de checkpoint y restart en clusters. *XXIV Congreso Argentino de Ciencias de la Computación (CACIC 2018)*, 2018.
- [16] Marina Morán, Javier Ballardini, Dolores Rexachs, and Enzo Rucci. Towards management of energy consumption in hpc systems with fault tolerance. In *2020 IEEE Congreso Biental de Argentina (ARGENCON)*, pages 1–8, 2020.