

Cómputo paralelo y distribuido: métricas de rendimiento, aplicaciones de datos masivos e inteligencia artificial

Javier Balladini¹, Marina Morán¹, Claudio Zanellato¹, Claudia Rozas¹, Rodrigo Cañibano¹, Darío Semenzato¹, Agustín Chiarotto¹, Cristina Orlandi², Armando De Giusti³, Remo Suppi⁴, Dolores Rexachs⁴, Emilio Luque⁴

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

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 la métrica del consumo energético, un tema de enorme relevancia actual dado el gran número de unidades de procesamiento que componen los sistemas. Por otro lado, buscamos aplicar las técnicas de cómputo paralelo y distribuido para solucionar problemas del sector salud. En particular, nos orientamos a soluciones de alertas tempranas de gravedad para Unidades de Cuidados Intensivos, afectadas por un gran volumen de datos y la necesidad de aplicar técnicas de inteligencia artificial. También nos enfocamos en una aplicación para atender y clasificar pacientes con COVID-19 según el riesgo de salud. Los trabajos se desarrollan en colaboración con otras universidades, dos hospitales públicos de Argentina y un centro de investigación. La formación de recursos humanos en estas líneas está orientada al nivel de licenciatura, grado, maestría y doctoral.

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

1. Contexto

Nuestra investigación es parte de una solicitud de nuevo proyecto de investigación de la Universidad Nacional del Comahue que está en evaluación, denominado "Cómputo paralelo y distribuido: métricas de rendimiento, aplicaciones de big data e inteligencia artificial". La mayoría de los temas surgieron en proyectos anteriores acreditados. El eje de aplicaciones para la salud se desarrolla en colaboración con el Hospital Francisco López Lima de General Roca (Río Negro), el Hospital Italiano y el Instituto de Medicina Traslacional e Ingeniería Biomédica de Buenos Aires. Respecto al eje relacionado con el consumo energético 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, y el grupo de investigación "High Performance Computing for Efficient Applications and Simulation" de la Universidad Autónoma de Barcelona, España.

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. Nos enfocamos principalmente en el sector de salud. A su vez, una de las mayores problemáticas que afectan a los sistemas de cómputo paralelo (y más aún a los de grandes dimensiones) es el elevado consumo energético y potencia eléctrica que demandan. Desde hace años no solo interesan las métricas de rendimiento computacional sino también energéticas. Por esta razón, también orientamos nuestra investigación 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 que direccionan ambas temáticas 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 (como monitores médicos de signos vitales). 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 (ubicados a los lados de las camas) a intervalos de tiempo (por ejemplo de dos horas o fracciones) que definen los médicos para cada paciente. El equipamiento médico emite alertas cuando hay riesgo en la salud del paciente basándose en mediciones de ciertos parámetros. Luego, los médicos analizan los datos de los formularios y dan a los enfermeros indicaciones de tratamientos a realizar.

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 recursos humanos y materiales. Esta situación es agravada por déficit de personal intensivista, causada porque es una profesión poco atractiva. Así, 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. Esta enfermedad es causada por un virus extremadamente contagioso. 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] durante la emergencia sanitaria de pacientes con COVID-19 por medio de la adaptación de uno de los sistemas de alertas tempranas más

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

conocido y validados. Esto permite mejorar el rendimiento del personal, reduciendo los controles en pacientes estables y 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. Así, nos hemos propuesto desarrollar un sistema informático inteligente que implemente un sistema de alerta temprana para pacientes en internación general con COVID-19, que realice 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 materiales y 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 para COVID-19 o similares en funcionalidad.

2.2. Consumo energético de sistemas de 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 415,5 PFlops frente a los 148,6 PFlops de su predecesor, Summit. Al mismo tiempo, estas supercomputadoras se encuentran entre las diez primeras de las más eficientes energéticamente del ranking Green500, con alrededor de 14,7 GFlops/W. Sin embargo, el consumo de energía sigue aumentando; mientras que la computadora Summit tiene un consumo de 10 MW, Fugaku sube a 28 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, 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, la tolerancia a fallos y el consumo energético se han identificado como los dos mayores desafíos a los que deberemos enfrentarnos [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 reducción del consumo energético de sistemas de HPC, y en la resolución de problemas que tengan una alta demanda computacional e impacto social en el campo de la salud.

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 del tipo smartphones y tablets.

3.2. Consumo energético de sistemas de HPC

Nos centramos en el desarrollo de metodologías, modelos y construcción de software para administrar y gestionar el consumo de energía y prestaciones 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, y el desarrollo de un dispositivo embebido adquirente de datos que envía datos a la plataforma de procesamiento por medio de una red WiFi. 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.

COVINDEX

Se construyó un modelo de alertas tempranas basado en el conocimiento de médicos expertos nacionales e internacionales. Se está desarrollando un sistema informático, denominado COVINDEX, que implementa dicho modelo. El sistema se compone de una aplicación web, una aplicación móvil y una aplicación que ejecuta en un servidor en la nube. El sistema se pondrá a disposición de los hospitales en Abril de este año. A partir de los datos históricos que se vayan recolectando de los 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.

4.2. Consumo energético de sistemas de 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.

5. Formación de recursos humanos

El equipo de trabajo de la Universidad Nacional del Comahue tiene un Doctor y un Magíster, una estudiante de Doctorado (a graduarse en el presente año) y dos estudiantes de Maestría (a graduarse en 2022). Tres estudiantes de grado están realizando trabajos de tesis, y hay un estudiante de tecnicatura. En 2020 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 Balladini, 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 Balladini, 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 Balladini, 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 Balladini, 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 Héroult, 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 Balladini. 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. Balladini, D. Rexachs, and E. Luque. Prediction of energy consumption by checkpoint/restart in hpc. *IEEE Access*, 7:71791–71803, 2019.
- [15] Marina Morán, Javier Balladini, 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 Balladini, Dolores Rexachs, and Enzo Rucci. Towards management of energy consumption in hpc systems with fault tolerance. *arXiv preprint arXiv:2012.11396*, 2020.