



M.Sc. Abel Brenes Arce, Centro de Informática

Clúster institucional HPC UCR: Liderando el futuro del procesamiento científico

5 SEPT 2024 Ciencia y Tecnología

En la era de los datos masivos, surge la necesidad de analizar y producir resultados certeros en el menor tiempo posible; esto ha impulsado a la industria hacia la búsqueda de esquemas óptimos de procesamiento de información. Un computador que actúa como un conjunto de recursos sincronizados, aunque podría ser robusto y especializado no cumpliría las demandas impuestas por tecnologías de alto procesamiento como el Big Data, Machine Learning, Redes neuronales, Inteligencia Artificial, entre otras.

Un Clúster de Computación de Alto Rendimiento (HPC: High Performance Computing, por sus siglas en inglés) es un conglomerado de recursos enfocado en el procesamiento distribuido de grandes volúmenes de datos y cálculos complejos a través de los recursos en diferentes nodos a la vez, mejorando en varios órdenes de magnitud la respuesta esperada, de días a minutos se tendrían los resultados en comparación con sistemas aislados. ¿Serían estos conglomerados de recursos computacionales como el Clúster HPC, la solución para atender necesidades de procesamiento de cálculos complejos con datos masivos, que demandan resultados en cada vez menos tiempo?

La escala para medir la potencia de cálculo de un sistema se lleva a cabo en operaciones de coma flotante por segundo (FLOPS: Floating-Point Operations, por sus siglas en inglés), la cual, se incrementa en un factor mil respecto a la escala anterior. Para ilustrar el rendimiento y el impacto de estos sistemas, un clúster que se desempeña en el orden de TeraFlops (un billón de operaciones por segundo, (o.p.s.) 10^{12}) se utilizan en áreas como la modelización meteorológica, la investigación en biología molecular, el diseño de medicamentos y la física de partículas. Un clúster con un rendimiento en PetaFlops (mil billones de o.p.s, 10^{15}) permitiría simulaciones complejas de fenómenos naturales, pronósticos climáticos a largo plazo, análisis genómicos a gran escala y aplicaciones avanzadas de aprendizaje automático. Con un ExaFlops (un trillón de o.p.s, 10^{18}) prepara el

terreno para modelos aún más complejos, como la simulación detallada del comportamiento de los materiales, la investigación de la fusión nuclear o la comprensión profunda de sistemas biológicos complejos. Sin embargo, la industria de semiconductores tiene retos y demandas insatisfechas respecto al crecimiento de las capacidades de los procesadores, cuya duplicación cada seis meses quedó atrás, estando a las puertas del límite de rendimiento.

El uso global del clúster HPC está abriendo las fronteras de investigaciones inimaginables que han servido, por ejemplo, para resolver aplicaciones prácticas, tales como el estudio y toma de decisiones sobre el comportamiento de la población que sirvieron para detener la propagación del COVID19. El Centro Europeo para la Investigación Nuclear (CERN) realiza experimentos sobre el choque de partículas que generan un petabytes de información por segundo, datos que se analizan para estudios sobre el posible origen del universo.

El Centro de Informática (CI) de la Universidad de Costa Rica promovió la contratación del Clúster Institucional de procesamiento de datos científicos en modalidad de servicios administrados (Concurso 2022LN-000015-00), a través de un esfuerzo institucional que culmina con la puesta en marcha del servicio de Clúster Institucional HPC UCR que representa el próximo salto generacional, consolidando y sobrepasando por mucho las capacidades de nueve clústeres predecesores sumados, que poco a poco están dejando de operar. La nueva plataforma se encuentra hospedada en dos bastidores en el Centro de Datos Institucional (CDI) administrado por el CI; con alta disponibilidad óptica, eléctrica y de climatización. Los clústeres HPC estimulan la construcción de redes colaborativas humanas. La operación de la infraestructura como servicio (IaaS, por sus siglas en inglés) del Clúster institucional está a cargo del proveedor del servicio administrado y del fabricante, mientras que la administración del servicio es responsabilidad de investigadores pioneros en HPC y del personal CI.

La ejecución de tareas en el clúster HPC se lleva a cabo en el elemento de procesamiento llamado núcleo (en inglés Core que corre un conjunto de instrucciones de forma independiente. El Core maneja hilos (Threads lo que permite que un núcleo trabaje en más de una tarea a la vez. Si tenemos un procesador con 4 núcleos soporta 2 hilos y puede manejar hasta 8 tareas simultáneamente. El núcleo forma parte del procesador, sea de uso general (CPU: Central Processing Unit, por sus siglas en inglés) o de gráficos (GPU: Graphics Processing Unit, por sus siglas en inglés). Un clúster HPC es una plataforma que se compone de nodos colaborativos interconectados por una red de datos y una red de almacenamiento de altas velocidades que transversalizan la información de los recursos de procesamiento, memoria y almacenamiento.

El Clúster Institucional HPC UCR es una arquitectura de nodos distribuidos que podríamos imaginar como un árbol invertido, inicia con el nodo de acceso (raíz) que verifica las credenciales de acceso a los servicios, los nodos maestros redundantes (ramas) que controlan la disponibilidad, asignación y ejecución de las tareas, con un grupo de nodos de procesamiento general (CPU), la rama de nodos densos con más de mil cores para cálculos y simulaciones y, finalmente, la rama de nodos especializados en procesamiento gráfico (GPU) con más de cincuenta y cinco mil núcleos. Todos los nodos corren el sistema operativo Linux adaptado a HPC. La ejecución de tareas es administrada por el gestor de recursos y colas SLURM (SLURM: Simple Linux Utility for Resource Management, por sus siglas en inglés), apoyado de un conjunto de bibliotecas para controlar el paralelismo, rutinas matemáticas, procesamiento gráfico, compiladores, la administración de archivos distribuidos, el acceso remoto y utilitarios como Ganglia para monitoreo en tiempo real. Para más información visite el sitio <https://hpc.ucr.ac.cr/>.

El administrador del clúster HPC se encarga de gestionar las colas de procesos, asignar recursos para la ejecución de programas que se distribuyen en piezas de código entre los múltiples nodos, en ejecución paralela o serial, para obtener los resultados de forma oportuna y rápida, basado en un conjunto de instrucciones (código) de propósito

particular, predecible y recurrente. El desarrollo de la Arquitectura de dispositivo unificada de computación (CUDA Compute Unified Device Architecture, por sus siglas en inglés) por la empresa NVIDIA, el Cuda Core se vuelve un núcleo de proceso más allá de los gráficos con la capacidad de procesamiento de propósito general potenciando aplicaciones de cómputo intensivo. Los nuevos GPU son catalizadores y potenciadores para el aprendizaje profundo y la Inteligencia Artificial (IA) al permitir entrenar modelos complejos de forma eficiente. El entrenamiento de sistemas de aprendizaje en clúster HPC, a partir de enormes volúmenes de datos y millones de fuentes de información, brindaría interacciones no programadas y resultados en lenguaje natural, un plus del sistema.

La gestión del Clúster institucional, sofisticado instrumento, requiere de un Modelo de Gobernanza que facilite el logro de objetivos como la universalidad, innovación, sostenibilidad y participación de áreas de investigación, docencia, acción social entre otros. La Gobernanza debe ser un marco claro y funcional liderado por científicos destacados en HPC, apoyados por ingenieros especializados, tecnología, infraestructura de avanzada como el CDI. La propuesta de estructura de Gobernanza del Clúster Institucional HPC UCR consiste en un Consejo Director pivote, responsable de tomar decisiones sobre el Clúster Institucional, las políticas de uso adecuado, la selección transparente de proyectos, la búsqueda de recursos y de un tejido de alianzas estratégicas que garanticen el potenciamiento de la plataforma, la venta de servicios y la permanencia en el tiempo. Un Equipo Consultivo conformado por la representación de distintas áreas del conocimiento con comunicación directa con el Consejo Directivo. Un Equipo Técnico encargado de ofrecer mantenimiento, asistencia y soluciones técnicas. Las Personas Usuarias representan el conjunto de personas con proyectos activos en el clúster. El Centro de Informática como instancia estratégica, asesora y técnica que administra recursos críticos para asegurar los objetivos de la Universidad de Costa Rica hacia una posición de vanguardia y excelencia.

El Clúster Institucional HPC UCR presenta además de un cambio generacional, un nuevo paradigma de procesamiento científico avanzado, por sus prestaciones y capacidades, construye un conglomerado de recursos que atiende las necesidades presentes y de corto plazo en procesamiento de cálculos complejos con datos masivos y soporte a la inteligencia artificial, ubicándose como número uno en la Universidad de Costa Rica, la región centroamericana y que destaca a nivel mundial.

¿Desea enviar sus artículos a este espacio?



Los artículos de opinión de *Voz experta UCR* tocan temas de coyuntura en textos de 6 000 a 8 000 caracteres con espacios. La persona autora debe estar activa en su respectiva unidad académica, facilitar su correo institucional y una línea de descripción de sus atestados. Los textos deben dirigirse al correo de la persona de la Sección de Prensa a cargo de cada unidad. En el siguiente enlace, puede consultar los correos electrónicos del personal en periodismo: <https://odi.ucr.ac.cr/prensa.html>

[M.Sc. Abel Brenes Arce, CISO](#)

abel.brenes@ucr.ac.cr

Etiquetas: [cluster](#), [investigacion](#), [tecnologia](#), [almacenamiento](#), [datos](#), [#vozexperta](#), [voz experta](#), [ucr](#).