
Cloud, HPC & Bio

Laboratorio 5: Explotando el paralelismo con Python en CPUs

Informe de valoración de prácticas.

En esta práctica se abordó el uso del paralelismo en Python aplicado a aplicaciones de cálculo científico intensivo en CPU, haciendo uso de los paquetes multiprocessing y Numba. A través de distintos ejercicios, como el cálculo de la suma de números primos y operaciones de reducción sobre grandes volúmenes de datos, se analizó el impacto del uso de múltiples núcleos en el tiempo de ejecución. Junto a esto, se reforzó la ejecución automatizada de notebooks mediante el gestor de colas SLURM, permitiendo la parametrización de los experimentos y su ejecución en un entorno HPC.

Entre los aspectos positivos de la práctica, destaca la comprensión clara de cómo la paralelización de tareas puede mejorar significativamente el rendimiento frente a implementaciones secuenciales. La comparación entre distintas técnicas, como el uso de multiprocessing con Pool y la paralelización mediante Numba con prange, permitió valorar las ventajas y limitaciones de cada enfoque. Además, la práctica resultó útil para afianzar el uso de herramientas habituales en entornos HPC, como Git para el control de versiones y SLURM para la gestión de trabajos.

Como principal limitación, la práctica requiere un conocimiento previo del entorno HPC para ejecutar correctamente los scripts en el contexto adecuado. En particular, la necesidad de lanzar los trabajos desde el nodo correcto puede no resultar evidente inicialmente, lo que puede provocar errores en la ejecución si no se sigue cuidadosamente el procedimiento establecido.