

Práctica 3 de Paralelismo

Carlos Torres
Daniel Sergio Vega

La práctica 3 consiste en paralelizar un código que calcula las distancias entre dos cadenas de ADN. Las bases de ADN están codificadas como enteros, y cada una de las cadenas se almacena en una matriz de M secuencias de ADN x N bases cada secuencia.

La descomposición de tareas que se hace es de dominio, dividiendo por filas de la matriz. Siendo p el número de procesos, a cada proceso se le asignan M/p (redondeado hacia arriba) filas de la matriz, y al último proceso se le asignan las filas restantes. En este caso se añade un padding al final del conjunto de datos para que todos los procesos reciban la misma cantidad de datos “brutos”, aunque el último proceso realmente calcula menos filas que los demás. Esto no es un problema porque, en el peor de los casos, el último proceso estaría calculando solamente una fila, pero el tiempo paralelo total seguiría siendo el mismo, ya que estaría determinado por el proceso que más carga computacional tiene.

La comunicación entre procesos se hace con dos operaciones Scatter para distribuir las dos cadenas iniciales, y una Gather para recuperar el vector de resultados. Además, se usan $2p$ operaciones Send y Recv para enviarle al proceso 0 los tiempos de los demás procesos.

Toda la salida por pantalla la realiza el proceso raíz (el 0 en nuestro caso). No hay entrada del usuario por teclado en ningún momento. La salida muestra los resultados de las distancias entre cada dos secuencias (sólo si está activado el flag de DEBUG), y muestra el tiempo que ha gastado cada proceso tanto en procesar su parte de los datos como en comunicaciones con los demás procesos.

Carlos Torres