

Abstrak

Multiple sequence alignment (MSA) adalah sebuah proses menyusun beberapa rantai protein menjadi satu konsensus baru rantai protein. Proses penyusunan rantai protein yang dilakukan oleh MSA menjadi tahapan awal dalam dunia *bioinformatics* untuk menghasilkan informasi biologis yang lebih banyak dibandingkan dengan metode *pairwise alignment*. Salah satu pendekatan yang digunakan dalam MSA adalah *heuristic approach* dengan aplikasi yang sering digunakan yaitu clustal yang menyusun panjang rantai protein sebanyak n secara sekuensial sehingga cukup memakan waktu ketika proses dilakukan. Hadoop sebagai lingkungan komputasi data besar memiliki kemampuan untuk memproses data yang berukuran besar menjadi lebih cepat karena adanya pemecahan tugas komputasi dan pemrosesan data menjadi potongan - potongan kecil yang diproses secara bersamaan oleh beberapa *client*. Dengan melihat keunggulan yang dimiliki oleh Hadoop tersebut, kami mentransformasikan model komputasi MSA yang diproses secara sekuensial menjadi sebuah model komputasi MSA baru yang dapat dijalankan di dalam lingkungan Hadoop. Dari hasil penelitian, kami mendapatkan hasil bahwa kecepatan rata – rata proses mapping 33,86 karakter/detik , kecepatan rata – rata proses reducing 12,8 karakter/detik, kecepatan proses keseluruhan sistem sebesar 5,49 karakter/detik, dan kecepatan proses keseluruhan model komputasi clustalW dalam lingkungan hadoop lebih lambat 722441,035 kali dibandingkan dengan kecepatan komputasi clustalW tanpa menggunakan lingkungan hadoop.

Kata kunci : *Multiple Sequence Alignment (MSA), Clustal, Hadoop.*