

ABSTRAK

Clustering merupakan sebuah teknik yang banyak digunakan untuk pendistribusian dan pengolahan data. Tujuan dari klustering itu sendiri adalah untuk menemukan struktur dasar dari sebuah data dan mengelompokkannya menjadi sekumpulan data yang mempunyai nilai untuk dapat dipelajari dan dianalisis lebih lanjut. Sebuah teknik pengelompokan dan pendistribusian data yang banyak digunakan saat ini adalah *K-Means Clustering*. *K-Means Clustering* banyak digunakan karena kemudahan dalam pengaplikasiannya serta memberikan hasil klustering yang cukup baik.

Ditengah era *Big Data* yang semakin berkembang seperti saat ini, penggunaan teknik dan analisis data yang masih bersifat tradisional ataupun serial mungkin tidak akan efisien lagi dalam pengolahan data yang jumlah dan ukurannya sangat besar. Maka dari itu penggunaan sebuah hardware ataupun system seperti Hadoop akan sangat membantu dalam proses klustering data yang sangat besar tersebut. Hadoop dapat digunakan secara efisien untuk pengolahan data dalam jumlah besar dikarenakan Hadoop memiliki sebuah algoritma pemrosesan data sendiri yang disebut *MapReduce*. *MapReduce* adalah sebuah algoritma yang dapat digunakan untuk mengatasi ukuran dan jumlah data yang besar dengan melakukan pendistribusian dan pengolahan data secara bersamaan.

Pada penelitian ini akan dianalisis bagaimana implementasi penggunaan *MapReduce* pada algoritma *K-Means Clustering* dengan menggunakan sebuah *Single Node* Hadoop yang akan dibandingkan dengan pemrosesan algoritma *K-Means Clustering* secara sekuensial dengan melihat waktu komputasinya.

Kata Kunci : Hadoop , *MapReduce*, *K-Means*, *Mapreduce K-Means*