

Abstrak

Implementasi Hadoop cluster untuk pengolahan data secara terdistribusi dalam skala besar sudah menjadi tren saat ini. Hadirnya hadoop cluster sangat membantu dalam bidang pengolahan data, banyak perusahaan yang mengimplementasikan hadoop cluster seperti facebook, yahoo, dan amazon. Hal ini didasari oleh kelebihan hadoop yang dapat memiliki performansi tinggi dengan menggunakan hardware sederhana.

Pada dasarnya pengimplementasian Hadoop cluster didasari pada kecepatan pengolahan data skala besar dengan menggunakan *hardware* yang standar namun dapat memiliki performansi yang baik.

Tujuan dari penelitian ini adalah mengimplementasikan hadoop cluster dengan menggunakan benchmark *wordcount* sebagai tools untuk mengetahui tingkat performansi dari jenis *text file* dengan memperhatikan kapasitas *Random Access Memory (RAM)*. Jenis-jenis *file text* yang akan diujicoba adalah doc, pdf, csv, xlsx dan txt.

Waktu ujicoba yang dihasilkan dari jenis-jenis *text file* tersebut menunjukkan urutan tingkat performansi terbaik dimulai dari jenis text file csv, txt, xlsx, pdf dan yang terakhir adalah jenis *file doc*. Waktu peningkatan performansi dari semua jenis *file* tidak mengalami peningkatan yang sama dengan peningkatan kapasitas RAM, pada saat kapasitas RAM ditingkatkan menjadi 100% hasil percobaan menunjukkan performansi dari jenis file doc mengalami peningkatan sebesar 4,58%, *file pdf* sebesar 7,57%, *file csv* sebesar 8,87%, file xlsx sebesar 8,35% dan *file txt* sebesar 12,82%.

Kata Kunci : Cluster, Hadoop, MapReduce, HDFS, RAM, Bandwidth