

Abstrak

Perancangan *Hadoop cluster* pada suatu jaringan virtual merupakan cara untuk menambah jumlah *node* yang berfungsi sebagai *tasktracker* dimana bertugas untuk melakukan proses pengolahan data pada suatu *cluster*. Hal tersebut bertujuan sebagai usaha untuk meningkatkan performansi *mapreduce* dimana secara teori semakin banyak *node* yang bertugas sebagai *tasktracker* maka proses yang dijalankan semakin cepat.

Pada tugas akhir ini dilakukan perancangan *Hadoop cluster* yang terdiri dari *Hadoop Distributed File System* (HDFS) sebagai sistem berkasnya dan *mapreduce* sebagai model pemrograman untuk pengolahan data. Penambahan *node* akan dilakukan pada tiga mesin fisik yang didalamnya terdapat maksimal empat mesin virtual. Pendekatan virtualisasi yang digunakan pada cluster yaitu *OS-level-virtualization* dengan menggunakan OpenVZ. *Cluster* yang berjalan pada lingkungan virtual tersebut akan dilakukan pengujian dengan melihat nilai *throughput* dari HDFS dan waktu eksekusi dari pengujian *sort*. Kemudian hasilnya akan dibandingkan dengan *mapreduce* yang hanya berjalan pada *node* fisik. Dari hasil pengujian dan analisa performansi, didapat bahwa cluster dengan skala fisik lebih baik apabila dibandingkan dengan skala virtual. Penambahan per tiga *node* virtual justru menaikkan rata-rata nilai waktu eksekusi sebesar sebesar 7.03% (3.69 detik).

Kata Kunci : cluster, hadoop, *mapreduce*, mesin virtual, virtualisasi.