

ABSTRAK

Saat ini teknologi server terus berkembang, demikian juga teknologi kluster komputer untuk berbagai keperluan dalam pengolahan data. Dengan hal ini penulis membuat komputer kluster serta mempelajari keunggulan komputer kluster untuk mengeksekusi algoritma *machine learning*. Dengan demikian dengan hadirnya komputer kluster ini akan menawarkan layanan untuk melakukan proses komputasi yang lebih cepat.

Untuk membuat komputer kluster diperlukan beberapa tahapan mulai dari instalasi Apache Hadoop sampai membagi node pada masing-masing komputer. Dilanjutkan dengan instalasi Apache Mahout sebagai perangkat lunak yang menyediakan algoritma *machine learning* untuk di eksekusi pada kluster komputer. Adapula Zabbix yaitu aplikasi monitoring yang digunakan untuk mengetahui sejauh mana kinerja CPU load dan trafik jaringan port ethernet yang digunakan selama proses eksekusi algoritma *machine learning* berlangsung.

Pada dataset 20 Newsgroups dengan ukuran file 36 MB, terlihat rata-rata lamanya waktu eksekusi dengan 1 node mencapai 22 menit 12 detik, 2 node mencapai 20 menit 42 detik dan 3 node hanya mencapai 18 menit 3 detik. Sedangkan pada dataset Wikipedia XML dengan ukuran file 1 GB, terlihat rata-rata lamanya waktu eksekusi dengan 1 node mencapai 22 menit 6 detik, 2 node mencapai 18 menit 39 detik dan 3 node mencapai 15 menit 6 detik. Hasil pengujian menunjukkan, ada perbandingan waktu yang cukup signifikan mencapai rata-rata 2 sampai 7 menit lebih cepat bila node ditambahkan lebih banyak, sehingga 3 node lebih cepat dari 2 node dan 1 node.

Kata Kunci: Komputer kluster, Apache Hadoop, Apache Mahout, *Machine learning*, Zabbix