

Analisis Performansi NoSQL *file system* pada *Hadoop Distribute File System* dan *Cassandra File System*

Juan Yudha¹, Sidik Prabowo², Siti Amatullah Karimah³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

⁴Divisi Digital Service PT Telekomunikasi Indonesia

¹juanyudha@students.telkomuniversity.ac.id, ²karimahsiti@telkomuniversity.ac.id,

³pakwowo@telkomuniversity.ac.id,

Abstrak

Pada saat ini *teknologi* berkembang dengan pesat. Berdasarkan penelitian yang dilakukan *Big Data* akan menguasai pasar teknologi. Tingkat pertumbuhan *Big Data* meningkat setiap tahunnya, dan *file system* (FS) yang dibutuhkan semakin besar. *Hadoop* merupakan OS *open source* yang sangat populer dikalangan *Big Data*. Namun dengan meningkatnya perkembangan dan penggunaan *Big Data*, spesifikasi pada *file system* juga mulai meningkat. Sistem yang ada hanya mampu untuk mentukan berdasarkan dari kategori *bottleneck* pada *Big Data*. Namun sistem ini dinilai masih kurang, karena tidak ada hasil keluaran dalam menentukan performansi dari suatu *file system*. Sehingga dibutuhkan sebuah sistem yang mampu untuk menentukan tigtat performansi dari suatu *file system*. *File System* saat ini menggunakan data yang tersusun secara terstruktur salah satu contohnya yaitu *SQL*. Untuk menangani data yang tidak terstruktur diperlukan suatu mekanisme *File System* yang menggunakan *NoSQL*. Untuk itu pada penelitian ini sudah di lakukan pengujian pada *file system* berdasarkan *load-write* dan *load-read* dengan menggunakan *Hadoop Distribute File System* (HDFS) dan *Cassandra File System* (CFS). Klasifikasi hasil dibagi menjadi dua yakni, bagus atau tidak nya dari sebuah *file system*. Dari penelitian yang sudah dilakukan di dapat kan hasil berupa bahwa *cassandra* memiliki performansi *throughput* yang lebih bagus dibandingkan dengan HDFS selisihnya yakni sebesar 1818.75 detik untuk setiap operasi yang dijalankan, dan untuk kecepatan dalam memproses data didapatkan hasil berupa bahwa *cassandra* lebih bagus dibandingkan dengan HDFS selisihnya yakni sebesar 6028 detik.

Kata kunci : *Big Data*, *performance*, *HDFS*, *CFS*, *load-read*, *load-write* merupakan kata-kata kunci yang menjelaskan isi tulisan, biasanya bisa diambil dari judul dan abstrak. Maksimal enam buah dan ditulis dengan huruf kecil, kecuali singkatan