

Abstrak

Hadoop adalah *platform open source* berbasis Java yang berada di bawah lisensi Apache dan digunakan untuk mendukung aplikasi yang berjalan pada *big data*. HDFS merupakan komponen dari Hadoop yang dapat menyimpan *file* yang sangat besar dengan akses data *streaming* dan berjalan pada kelompok komoditas perangkat keras. HDFS dirancang untuk menangani *file* besar dalam jumlah banyak yang ukuran *file* nya hingga petabyte, exabyte. Di era sekarang ini, *Big Data* menjadi topik yang menarik untuk dibahas dengan bentuk data yang berbagai *variety, velocity* dan *volume* data. Di dalam *Big Data* terdapat *file-file* yang berukuran besar dan kecil yang akan di proses HDFS. Akan tetapi terdapat masalah yang di temukan ketika HDFS menangani *file* kecil dalam jumlah banyak, sehingga terdapat beberapa solusi yang ditawarkan untuk menangani *file* kecil dalam jumlah banyak di HDFS yaitu dengan menggunakan metode HAR (*Hadoop Archive*) dan *Combine File Input Format*. Dengan kedua metode ini maka *file* kecil di HDFS teratasi. kemudian di lakukan perbandingan yang mana metode yang digunakan menghasilkan penggunaan block dan waktu pemrosesan yang di gunakan sedikit.

Kata kunci : Hadoop, HDFS, HAR, combine file input format, file kecil, block.