

ABSTRAK

Teknologi *Big data* merupakan kumpulan data dalam skala besar, yang mempunyai karakteristik data yang variatif, sangat cepat pertumbuhannya dan kompleks datanya. Data yang kompleks merupakan data yang tidak terstruktur yang perlu diolah khusus dengan suatu infrastruktur yang dapat mengelola data dalam volume besar berjalan secara realtime.

Untuk itu diperlukan suatu metode yaitu Mapreduce, guna memudahkan komputasi yang akan dilakukan pada *big data*. Mapreduce digunakan untuk melakukan komputasi kumpulan data yang terdapat pada *Hadoop Distributed File System* (HDFS). Metode Mapreduce dapat ditransformasi dengan berbagai bentuk. Dengan Apache Flink metode Mapreduce dapat dikaji kembali pada arsitektur yang berbeda.

Pada tugas akhir ini pengelolaan data berupa data yang tidak terstruktur dalam bentuk teks. Merancang Aplikasi HDFS pada sistem operasi linux dan mengimplementasikan metode Mapreduce. Program mapreduce yaitu berupa program penghitung jumlah kata menggunakan fungsi yang terdapat pada Apache Flink. Pada penelitian ini, Flink Mapreduce dapat melakukan komputasi lebih cepat sekitar 37,18% dari Hadoop Mapreduce.

Keyword : Big Data, Hadoop Apache, MapReduce, Apache Flink