

ABSTRAK

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.

Pada tugas akhir ini digunakan metode MapReduce untuk memudahkan komputasi yang akan dilakukan pada suatu *big data*. MapReduce merupakan suatu model pemrograman untuk menulis aplikasi yang dapat memproses suatu *big data* secara paralel pada beberapa *node*. MapReduce memberikan kemampuan analitis untuk menganalisis *volume* besar data yang kompleks. *Platform* yang digunakan adalah Hadoop, Hadoop mempunyai algoritma MapReduce sendiri.

Tugas akhir ini akan menganalisis performa dari Hadoop MapReduce dan membandingkannya dengan Apache Spark yaitu *platform* yang dibuat untuk memproses suatu *big data* yang dikembangkan berdasarkan Hadoop MapReduce dengan peningkatan performa pemrosesan. Skenario yang digunakan adalah memproses *wordcount* suatu data dengan besar data yang berbeda yang bertujuan untuk menganalisis *response time* dan penggunaan *hardware* dari kedua *platform* tersebut.

Kata kunci: *Big Data, Apache Spark, MapReduce, Hadoop*