

INTISARI

ANALISIS PERFORMA DISTRIBUSI DATA PADA SISTEM TERDISTRIBUSI DENGAN HADOOP MAPREDUCE

Oleh

IMAM FACHRUL RAZI

20/466414/PPA/05980

Perkembangan dunia teknologi semakin pesat dimana kecepatan pertumbuhan data juga menjadi sangat cepat dan menjadi penting. Tren menunjukkan bahwa pertumbuhan data akan terus berlipat ganda dari waktu ke waktu, sehingga pada akhirnya akan melampaui batas penyimpanan dan sistem basis data yang tersedia. Hal tersebut memungkinkan untuk membuat metode khusus untuk menangani pemrosesan *big data* secara efektif dan efisien dalam sebuah *cluster database*.

Penerapan skema *multiple node* merupakan salah satu cara dalam proses pendistribusian data yang memiliki jenis variasi dan ukuran besar karena akan berdampak pada peningkatan kinerja proses dan juga sistem pengolahan data akan menjadi lebih mudah dan proses pendistribusian data akan membagi sejumlah besar data ke beberapa *node* atau media penyimpanan yang berjalan dalam sebuah *cluster* untuk diproses.

Dalam penelitian ini mengusulkan rancangan skema distribusi data pada sistem terdistribusi yang akan diterapkan pada *physical* dan *virtual environment* menggunakan *Hadoop MapReduce* sebagai salah satu *framework big data* dalam bentuk *cluster multiple node* yang terdiri dari satu *master node* dan dua *slave node* untuk fungsi pendistribusian dan penyimpanan data kemudian menganalisis hasil dari pengujian kinerja proses distribusi menggunakan metode yang ada.

Kata Kunci: *Data Distribution, Physical Environment, Virtual Environment, Hadoop, MapReduce, Multiple Node.*

ABSTRACT

ANALYSIS OF DATA DISTRIBUTION PERFORMANCE ON A DISTRIBUTED SYSTEM WITH HADOOP MAPREDUCE

By

IMAM FACHRUL RAZI

20/466414/PPA/05980

The development of the world of technology is increasingly rapid where the speed of data growth is also very fast and becomes important right now. Trends show that data growth will continue to multiply over time, so that it will eventually exceed the limits of available storage and database systems. This makes it possible to create some custom methods to handle big data processing effectively and efficiently in a database cluster.

The application of a multiple node scheme is one way in the data distribution process that has a large variety and size because it will have an impact on improving process performance and also the data processing system will become easier. The data distribution process will divide large amounts of data into several nodes or different storage media running in a cluster for processing.

In this study, we propose a data distribution scheme design for a distributed system that will be applied to physical and virtual environments using Hadoop MapReduce as one of the big data frameworks in the form of multiple node clusters consisting of one master node and two slave nodes for the function of distributing and storing data later. analyze the results of testing the performance of the distribution process using existing methods.

Key Word: Data Distribution, Physical Environment, Virtual Environment, Hadoop, MapReduce, Multiple Node.