

INTISARI

SISTEM OPERASI UNTUK PEMROSESAN BIG DATA DENGAN BERBASIS CENTOS 7

Carolus Gaza Nindra Tama
13/347460/PA/15250

Big data adalah kondisi dimana model penyimpanan basis data konvensional tidak dapat lagi menanggulangi data dengan jumlah yang besar. *Big data* memiliki karakteristik lima V, yaitu *Volume*, *Velocity*, *Variety*, *Value*, dan *Veracity*. Beberapa tahun terakhir ini, sudah terdapat banyak sekali *framework-framework* yang dapat digunakan untuk pemrosesan *big data*. Framework tersebut dapat diinstal dalam sistem operasi yang telah tersedia saat ini, seperti Centos dan Ubuntu, tetapi sistem operasi tersebut masih belum optimal untuk pemrosesan *big data*.

Penelitian ini bertujuan untuk membuat suatu sistem operasi yang memiliki kemampuan untuk pemrosesan *big data*. Pada penelitian ini, dilakukan modifikasi kernel linux untuk meningkatkan performa pada pemrosesan *big data* dan juga penggabungan beberapa *framework big data* ke dalam sebuah sistem operasi. Metode yang digunakan untuk pembuatan sistem operasi ini adalah metode *remastering* dengan menggunakan Centos 7 sebagai sistem operasi dasar. Sistem operasi hasil *remastering* ini diberi nama Gamabox OS.

Berdasarkan hasil pengujian yang telah dilakukan, kernel Gamabox OS mendapatkan peningkatan performa 12% lebih baik dibandingkan dengan kernel Centos 7. Selain itu, *framework big data* dari Gamabox OS memiliki kinerja yang lebih efisien dan memiliki waktu eksekusi yang lebih kecil daripada HDP. Setelah diuji dalam kluster dengan dua *node*, performa dapat meningkat sebesar 32 % pada pengujian HPL dan sebesar 49% pada pengujian TeraSot dibandingkan dengan satu *node*.

Kata Kunci: *big data*, sistem operasi, *remastering*, linux, kernel, Centos 7

ABSTRACT

OPERATING SYSTEM FOR BIG DATA PROCESSING BASED ON CENTOS 7

Carolus Gaza Nindra Tama
13/347460/PA/15250

Big data is a condition where a conventional database storage model can not cope with large amounts of data. Big Data has the five V characteristic, namely Volume, Velocity, Variety, Value, and Veracity. Five V shows a problem and challenge which faced in big data. In recent years, there have been many frameworks that can be used for big data processing. The framework can be installed in currently available operating system, such as Ubuntu and Debian, but the operating system still not optimal for big data processing.

This research aims to create an operating system that has an ability to process big data. In this research, I modified the linux kernel to improve performance on big data processing and also merge some big data frameworks into an operating system. The method that be used to make this operating system is remastering by using Centos 7 as the base operating system. Remastered operating system is named Gamabox OS.

Based on the results of tests that have been done, the Gamabox OS kernel get 12% better performance improvement compared with Centos 7 kernel. In addition, big data framework from Gamabox OS has more efficient performance and less execution time than HDP. After being tested in a cluster, the performance can increase 32% in HPL testing and 49% in TeraSort testing compared with one node.

Keywords: big data, operating system, remastering, linux, kernel, Centos 7