

ABSTRAK

Pertumbuhan data ternyata sangat mempengaruhi perkembangan volume dan jenis data yang terus meningkat di dunia maya. Jenis data, mulai dari data yang berupa teks, gambar atau foto, video hingga bentuk data-data lainnya membanjiri sistem komputasi. Dengan meningkatnya volume data, pengolahan data perlu diubah dengan metode yang lebih efektif. Salah satu metode yang dapat mengolah data lebih baik adalah komputasi terdistribusi. Salah satu perangkat lunak yang telah menggunakan metode komputasi terdistribusi adalah Hadoop. Penelitian ini dilakukan untuk menguji karakteristik dari Hadoop dengan melakukan pengujian. Tiga macam skenario diimplementasikan untuk mengamati bagaimana karakteristik Hadoop dan menganalisa waktu penerimaan data. Pengujian dilakukan dengan menggunakan *file* berukuran 800MB, 1.2GB, dan 2GB. Berdasarkan penelitian, diketahui bahwa penambahan ukuran data, dari 800 MB hingga menjadi 2 GB dapat memperlambat rata-rata waktu penerimaan data pada penambahan ukuran data selama 14,147 detik pada *slave* 1 dan 9,516 detik pada *slave* 2. Penambahan *block size* juga memperlambat rata-rata waktu penerimaan data pada ukuran data 800 MB selama 19,340 detik.

Keyword : *big data*, distribusi, Hadoop, sistem *file*.



ABSTRACT

Data growth can affect the development of volume and data type. All data type such as text, picture, video and other data types is overwhelming the computing system. With the increasing of data volume, data processing method need to be changed into more effective method. One of the effective method is the distributed computing. Hadoop is one of softwares that used distributed computing method. The purpose of This study is to observe the characteristic of Hadoop. Three scenarios are implemented to observe the characteristic of Hadoop and to analyze the data receiving time. Each scenarios using three different data sizes: 800 MB, 1.2 GB and 2 GB. Based on the study, known that the additional of size with suitable specifications can decelerate the average of receiving time on each data size for 14,147 seconds on slave 1 and 9,516 seconds on slave 2. The additional of block size also can decelerate the average of receiving time on data with 800 MB size for 19,340 seconds. Based on the study also known that Hadoop is still replicate the data blocks with the designed interferences.

Keywords: big data, distributed data, file system, Hadoop.



DAFTAR ISI

| | |
|---|------|
| LEMBAR PENGESAHAN | i |
| PERNYATAAN ORISINALITAS LAPORAN PENELITIAN..... | ii |
| PERNYATAAN PUBLIKASI LAPORAN PENELITIAN | iii |
| PRAKATA..... | iv |
| ABSTRAK | vi |
| ABSTRACT..... | vii |
| DAFTAR ISI..... | viii |
| DAFTAR GAMBAR..... | x |
| DAFTAR TABEL..... | xi |
| BAB 1 PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 2 |
| 1.3 Tujuan Pembahasan | 2 |
| 1.4 Ruang Lingkup..... | 2 |
| 1.5 Sumber Data..... | 3 |
| 1.6 Sistematika Penyajian | 3 |
| BAB 2 KAJIAN TEORI | 5 |
| 2.1 Big Data | 5 |
| 2.2 Hadoop | 5 |
| 2.2.1 Hadoop Single-node..... | 6 |
| 2.2.2 Hadoop <i>Multi-node</i> | 6 |
| 2.2.3 Hadoop <i>Distributed File System</i> | 7 |
| 2.3 MapReduce | 9 |
| 2.4 Cacti | 10 |

| | |
|--|-----|
| 2.5 OpenSSH..... | 10 |
| 2.6 Yet Another Resource Negotiator (YARN)..... | 12 |
| 2.7 Penelitian Terkait | 13 |
| 2.7.1 Analisis Pengaruh Ukuran Data terhadap Waktu Penerimaan Data | 13 |
| 2.7.2 Analisis Pengaruh <i>Block size</i> Terhadap Waktu Penerimaan Data | 14 |
| BAB 3 ANALISIS DAN RANCANGAN SISTEM..... | 16 |
| 3.1 Perancangan | 16 |
| 3.2 Skenario Pertama | 16 |
| 3.3 Skenario Kedua..... | 17 |
| 3.4 Skenario Ketiga..... | 17 |
| BAB 4 IMPLEMENTASI..... | 19 |
| 4.1 Lookup Library | 19 |
| 4.2 Instalasi java..... | 19 |
| 4.3 Konfigurasi SSH | 19 |
| 4.4 Konfigurasi Hadoop..... | 20 |
| 4.5 Mengubah ukuran <i>block</i> | 21 |
| BAB 5 PENGUJIAN | 22 |
| 5.1 Pengujian Skenario Pertama | 22 |
| 5.2 Pengujian Skenario Kedua..... | 24 |
| 5.2.1 Hasil Pengujian | 24 |
| 5.3 Pengujian Skenario Ketiga..... | 26 |
| BAB 6 SIMPULAN DAN SARAN..... | 28 |
| 6.1 Simpulan | 28 |
| 6.2 Saran..... | 28 |
| DAFTAR PUSTAKA | 29 |
| LAMPIRAN A DATA PENGUKURAN | A-1 |

DAFTAR GAMBAR

| | |
|---|----|
| Gambar 2.1 Hadoop <i>single-node</i> | 6 |
| Gambar 2.2 <i>Multi-node cluster</i> | 7 |
| Gambar 2.3 Penulisan data pada HDFS | 7 |
| Gambar 2.4 <i>MapReduce logical data flow</i> | 10 |
| Gambar 2.5 Sequence diagram untuk pemisahan privilege OpenSSH | 11 |
| Gambar 2.6 Bagaimana YARN menjalankan aplikasi..... | 12 |
| Gambar 2.7 Grafik pengaruh <i>physical machine</i> terhadap kecepatan MapReduce | 14 |
| Gambar 2.8 Grafik efek dari HDFS <i>block size</i> pada ukuran <i>file</i> 512 MB | 15 |
| Gambar 3.1 Topologi pengujian | 16 |
| Gambar 3.2 Ilustrasi Skenario Kedua | 17 |
| Gambar 4.1 Konfigurasi <i>lookup</i> | 19 |
| Gambar 5.1 Pembagian <i>block</i> pada Web Service HDFS | 22 |
| Gambar 5.2 (a)Block ID pada slave 1 (b) <i>block</i> ID pada slave 2..... | 23 |
| Gambar 5.3 Grafik rata-rata waktu penerimaan data skenario pertama..... | 23 |
| Gambar 5.4 Log <i>Slave</i> 1 | 24 |
| Gambar 5.5 Menerima <i>block</i> setelah start semua daemons | 25 |
| Gambar 5.6 Block ID yang disimpan di direktori <i>rbw</i> pada slave 2..... | 26 |
| Gambar 5.7 Grafik rata-rata waktu penerimaan data pada slave 1 dan slave 2 .. | 26 |
| Gambar 5.8 Grafik rata-rata waktu penerimaan data skenario ketiga..... | 27 |

DAFTAR TABEL

| | |
|--|----|
| Tabel 2.1 Hasil kecepatan rata-rata MapReduce menggunakan <i>physical machine</i> | 13 |
| Tabel 2.2 Hasil percobaan dengan ukuran <i>file</i> 512 MB..... | 14 |
| Tabel 3.1 Spesifikasi mesin yang digunakan | 16 |
| Tabel 4.1 Konfigurasi core-site..... | 20 |
| Tabel 4.2 Konfigurasi hdfs-site..... | 20 |
| Tabel 4.3 Konfigurasi yarn-site..... | 20 |
| Tabel 4.4 Konfigurasi ukuran <i>block</i> | 21 |
| Tabel 5.1 Rata-rata waktu penerimaan data skenario pertama | 23 |
| Tabel 5.2 Rata-rata waktu penerimaan data skenario ketiga..... | 26 |

