

PENGEMBANGAN SISTEM BIG DATA: RANCANG BANGUN INFRASTRUKTUR DENGAN FRAMEWORK HADOOP

Ahmad Sudirman, Irawan, Zawiyah Saharuna

Teknik Informatika dan Komputer, Politeknik Negeri Ujung Pandang, Indonesia

Info Artikel

Riwayat artikel:

Received, (1 Juni 2024)

Revised, (15 Juni 2024)

Accepted, (22 Juni 2024)

Kata kunci:

HDFS,

Map Reduce,

Block Size.

ABSTRACT

Hadoop is a distributed data storage platform that provides a parallel processing framework. HDFS (Hadoop Distributed File System) and Map Reduce are two very important parts of Hadoop, HDFS is a distributed storage system built on java, while Map Reduce is a program for managing large data in parallel and distributed. The research focused on testing the data transfer speed on HDFS, testing using 3 types of data, namely video, ISO Image, and text with a total of 512 MB, 1 GB, and 2 GB of data. Testing will be carried out by entering data into HDFS using the Hadoop command and changing the size of the block size with parameters 128 MB, 256 MB and 384 MB. Hadoop's performance on a large block size of 384 MB has better speed than block sizes of 128 MB and 256 MB because the data will be divided into 384 MB so that data mapping will be less than on block sizes of 128 MB and 256 MB.

ABSTRAK

Hadoop adalah platform penyimpanan data terdistribusi yang menyediakan kerangka pemrosesan paralel. HDFS (Hadoop Distribud File System) dan Map Reduce merupakan dua bagian yang sangat penting pada Hadoop, HDFS adalah sistem penyimpanan terdistribusi yang dibangun diatas java, sedangkan Map Reduce adalah program untuk mengelolah data besar secara paralel dan terdistribusi. Pada penelitian difokuskan pada pengujian kecepatan transfer data pada HDFS, pengujian menggunakan 3 tipe data yaitu video, ISO Image, dan text dengan jumlah data 512 MB, 1 GB, dan 2 GB. Pengujian akan dilakukan dengan cara memasukkan data ke HDFS dengan menggunakan perintah Hadoop serta besar block size diubah dengan parameter 128 MB, 256 MB, dan 384 MB. Performance Hadoop pada besar block size 384 MB memiliki kecepatan yang lebih baik dari block size 128 MB dan 256 MB dikarenakan data akan terbagi menjadi 384 MB sehingga pemetaan data akan lebih sedikit dibanding pada block size 128 MB dan 256 MB.

Penulis Korespondensi:

Irawan

Teknik Informatika dan Komputer, Politeknik Negeri Ujung Pandang, Jl. Perintis Kemerdekaan KM. 10, Makassar, 90245

Email: irawan@poliupg.ac.id

1. PENDAHULUAN

Infrastruktur *Big Data* pada saat ini sangat penting dikarenakan dampak dari perkembangan teknologi yang signifikan, sehingga data elektronik semakin banyak digunakan oleh perusahaan yang dulunya hanya memiliki data yang berkisar *gigabyte* hingga *terabyte*, namun sekarang mencapai *pentabyte*. *International Data Corporation (IDC)* memprediksikan bahwa jumlah total data digital yang dibuat di seluruh dunia akan menjamur dari 4,4 *zettabyte* pada tahun 2013 menjadi 44 *zettabyte* pada tahun 2020 (IDC, 2016). Sekarang, *IDC* percaya bahwa pada tahun 2025 jumlahnya akan mencapai 180 *zettabyte*. Pertumbuhan yang mencengangkan berasal dari jumlah perangkat yang menghasilkan data serta jumlah sensor di setiap perangkat sekitar 11 miliar perangkat terhubung ke *internet* sekarang. Angka tersebut diperkirakan hampir tiga kali lipat menjadi 30 miliar pada 2020 dan kemudian hampir tiga kali lipat menjadi 80 miliar lima tahun kemudian [7].

Pada dasarnya untuk mengelolah data yang besar memerlukan aplikasi yang membutuhkan spesifikasi komputer dengan sumber daya yang tinggi. Biasanya harga komputer dengan sumber daya yang tinggi tidaklah murah, sedangkan untuk komputer dengan spesifikasi yang tidak terlalu tinggi akan kurang *reliable* dalam

menangani data yang begitu besar [14]. Dengan demikian dibutuhkan teknologi yang benar-benar dapat meningkatkan sebuah kinerja sistem.

Perlunya infrastruktur yang dapat menyimpan dan mengelolah data agar mudah diakses serta dikelola. Dalam hal ini sangat penting pengolahan data, sehingga dapat digunakan dalam pengambilan keputusan pada suatu organisasi tertentu baik itu *profit* maupun *non-profit*, baik itu swasta maupun pemerintah.

Hadoop menawarkan sistem terdistribusi yang bekerja pada komputasi paralel dikembangkan oleh *Apache Hadoop* karena terinspirasi oleh *Google File System (GFS)* dan *Map Reduce* yang dikembangkan oleh *Google inc* [16]. *Apache Hadoop* mengembangkan beberapa proyek diantaranya *Hadoop Distributed File System (HDFS)* dan *MapReduce Framework* untuk menyelesaikan permasalahan dengan melibatkan data yang besar yang berbasis *java* dan *open source* [16]. *Hadoop* memiliki beberapa kelebihan yaitu mengatasi overload jaringan, multitasking dilakukan secara independen, simpel program karena hanya menggunakan satu model programming yaitu *mapreduce*, *HDFS* sederhana dan dapat diandalkan dan *HDFS* dapat diakses oleh banyak client dengan menambahkan beberapa cluster server [4].

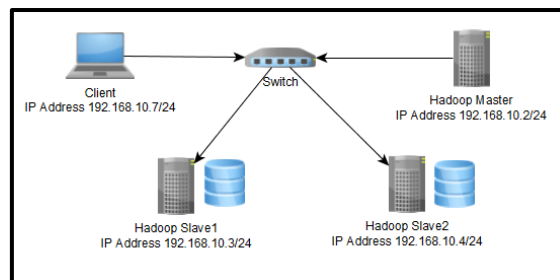
Berdasarkan permasalahan yang telah diuraikan di atas, judul yang diajukan pada penelitian ini adalah Rancang Bangun Infrastruktur *Big Data* Menggunakan *Hadoop Framework*.

2. METODE

2.1 Desain dan Perancangan

Pada penelitian ini diperlukan sebuah desain dan perancangan sistem sebagai acuan penelitian.

2.1.1 Desain Topologi *Hadoop Framework*



Gambar 1. Desain Topologi Jaringan Penelitian *Hadoop Framework*

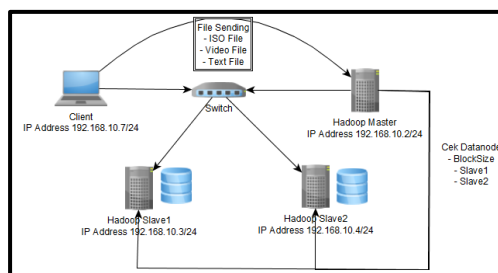
Dari hasil studi literatur diambil simpulan dari beberapa sumber untuk menentukan model sistem terbaik dan sesuai dengan penelitian. Adapun desain topologi jaringan yang akan diterapkan dalam penelitian ini adalah seperti pada Gambar 1.

Berdasarkan desain topologi jaringan penelitian *Hadoop Framework* Gambar 3.1 dapat dilihat bahwa :

1. *Master* merupakan komputer yang bertugas menyimpan informasi tentang distribusi *block-block* data pada *cluster hadoop*, dan mengatur pendistribusian *block-block* data tersebut di dalam *cluster hadoop* secara keseluruhan.
2. *Slave* merupakan komputer yang bertugas menyimpan potongan-potongan data yang didistribusikan kepadanya dan secara berkalah melaporkan kondisi dirinya ke pada *Namenode*.

2.2 Pengujian Sistem

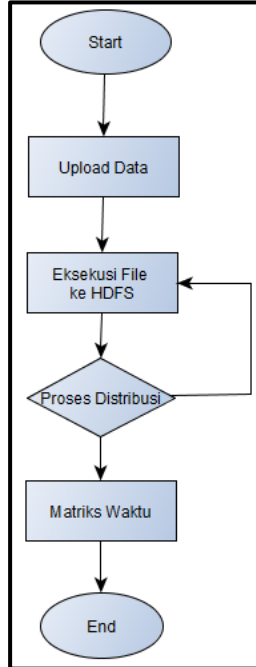
Tahapan pengujian sistem, pada tahapan ini akan dilakukan pengujian terhadap sistem yang telah dirancang dan diimplementasikan. Pengujian akan dilakukan untuk mengetahui kinerja dari *HDFS* dan *Map Reduce*. Pada Gambar 2 adalah skenario yang akan dilakukan pada pengujian ini :



Gambar 2. File Sending

1. Pada Gambar 2 adalah skenario mengunggah data dari *local* ke *master (HDFS)*, pada proses ini dapat dilihat sistem kerja dari *Namenode* ke *Datanode* serta *Job Tracker* ke *Task Tracker*.
2. Melihat Salah satu kinerja parameter dari *HDFS* yaitu kecepatan transfer data kedalam *block-block Datanode*. Pada Gambar 3.3 merupakan tahapan-tahapan yang akan dilakukan dalam pengujian kecepatan distribusi dari *hadoop*.

Gambar 3. Tahapan Pengujian



3. HASIL DAN PEMBAHASAN

3.1 Pengaruh *Block Size* Terhadap Kecepatan Distribusi

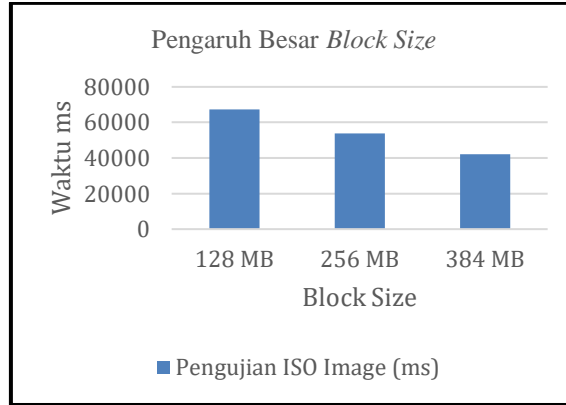
Pada Tabel 1, Tabel 2 dan Tabel 3 akan dijelaskan perbandingan kecepatan *HDFS* dalam mendistribusikan *File* yang di inputkan.

3.1.1 Pengaruh Besar *Block Size* Terhadap Kecepatan Distribusi Pada *File* 512 MB

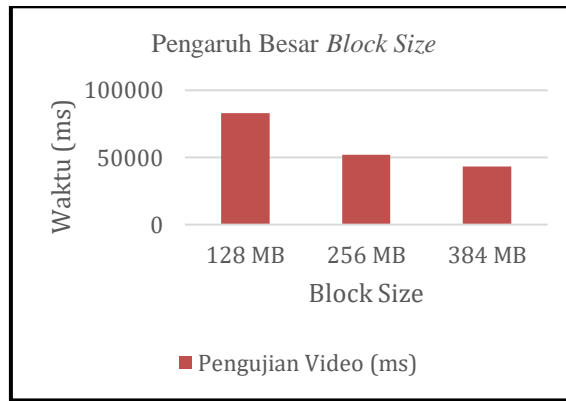
Tabel 1. Pengaruh Besar *Block Size* Terhadap kecepatan distribusi *file* 512 MB

Besar <i>Block Size</i>	Pengujian <i>Video</i> (ms)	Pengujian <i>ISO Image</i> (ms)	Pengujian <i>Text</i> (ms)
128 MB	82.890	67.250	79.549
256 MB	52.008	53.803	44.818
384 MB	43.089	42.049	34.554

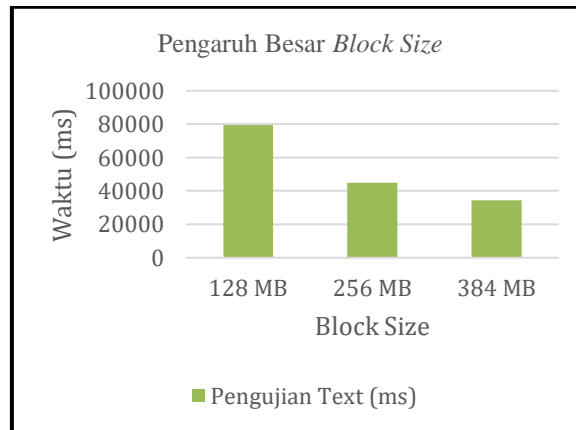
Dapat dilihat pada Tabel 4.1 kecepatan rata-rata dari pengimputan data baik *Video*, *ISO Image* maupun *Text* mengalami perubahan apabila parameter *block size* kita ubah yakni 128 MB menjadi 256 MB dan 384 MB dimana kecepatan pada saat ukuran *block size* 128 MB kecepatan rata-ratanya adalah 67.250 ms dibandingkan dengan pada ukuran *block size* 256 MB yakni 53.803 ms dan pada ukuran *block size* 384 MB 42.046 ms (*ISO Image*), sedangkan pada *file Video* dan *Text* dapat dilihat pada Tabel 4.1 dan Grafik perbandingan pada Gambar 4.2 dan Gambar 4.3, sehingga dapat disimpulkan bahwa semakin besar *Block Size* pada *HDFS* maka semakin cepat data didistribusikan. Berikut tampilan kecepatan data dalam bentuk grafik.



Gambar 4. Pengaruh Besar *Block Size* Terhadap kecepatan pada ukuran file 512 MB (*ISO Image*)



Gambar 5. Pengaruh Besar *Block Size* Terhadap kecepatan pada ukuran file 512 MB (*Video*)



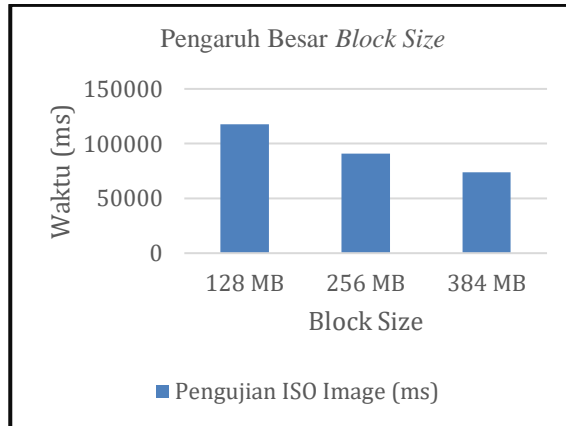
Gambar 6. Pengaruh Besar *Block Size* Terhadap kecepatan pada ukuran file 512 MB (*Text*)

3.1.2 Pengaruh Besar *Block Size* Terhadap Kecepatan Distribusi Pada File 1 GB

Tabel 2 Pengaruh Besar *Block Size* Terhadap kecepatan distribusi file 1 GB

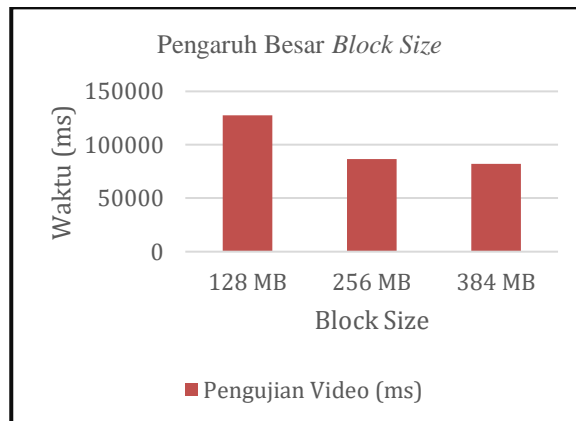
Besar <i>Block Size</i>	Pengujian <i>Video</i> (ms)	Pengujian <i>ISO Image</i> (ms)	Pengujian <i>Text</i> (ms)
128 MB	198.689	194.301	237.648
256 MB	180.077	188.805	130.573
384 MB	149.528	158.421	120.429

Dapat dilihat pada Tabel 2 kecepatan rata-rata dari pengimputan data baik *Video*, *ISO Image* maupun *Text* mengalami perubahan apabila parameter *block size* kita ubah yakni 128 MB menjadi 256 MB dan 384 MB dimana kecepatan pada saat ukuran *block size* 128 MB, kecepatan rata-ratanya adalah 117.517 ms dibandingkan dengan pada ukuran *block size* 256 MB yakni 90.702 ms dan pada ukuran *block size* 384 MB 73.981 ms (*ISO Image*), sedangkan pada *file Video* dan *Text* dapat dilihat pada Tabel 4.2.

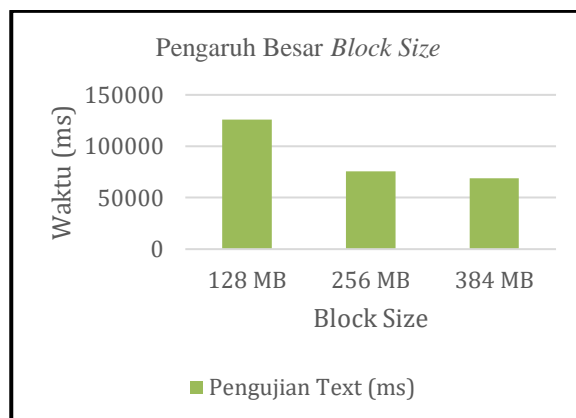


Gambar 7. Pengaruh Besar *Block Size* Terhadap kecepatan pada ukuran *file* 1 GB (*ISO Image*).

Dan Grafik perbandingan pada Gambar 8 dan Gambar 9, sehingga dapat disimpulkan bahwa semakin besar *block Size* pada *HDFS* maka semakin cepat data didistribusikan. Berikut tampilan kecepatan data dalam bentuk grafik.



Gambar 8. Pengaruh Besar *Block Size* Terhadap kecepatan pada ukuran *file* 1 GB (*Video*).



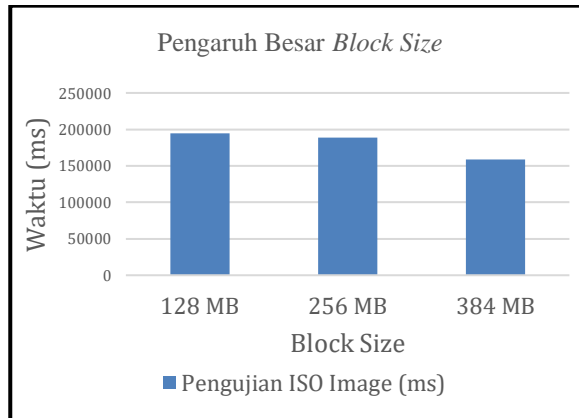
Gambar 9. Pengaruh Besar *Block Size* Terhadap kecepatan pada ukuran *file* 1 GB (*Text*)

3.2 Pengaruh Besar *Block Size* Terhadap Kecepatan Distribusi Pada *file 2 GB*

Tabel 3. Pengaruh Besar *Block Size* Terhadap kecepatan distribusi *file 2 GB*

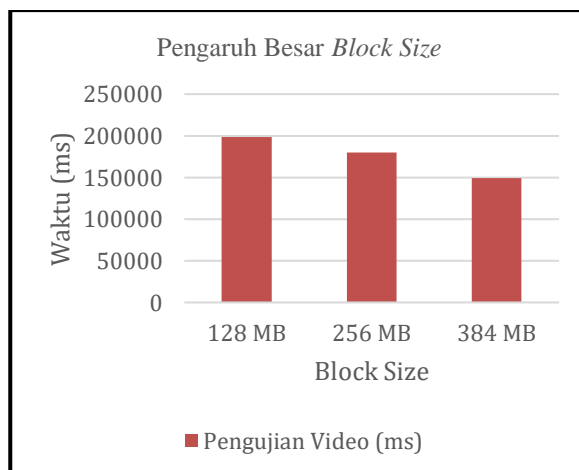
Besar <i>Block Size</i>	Pengujian <i>Video (ms)</i>	Pengujian <i>ISO Image (ms)</i>	Pengujian <i>Text (ms)</i>
128 MB	127.852	117.517	126.194
256 MB	86.901	90.702	75.807
384 MB	82.359	73.981	69.040

Dapat dilihat pada Tabel 3 kecepatan rata-rata dari pengimputan data baik *Video*, *ISO Image* maupun *Text* mengalami perubahan apabila parameter *block size* kita ubah yakni 128 MB menjadi 256 MB dan 384 MB dimana kecepatan pada saat ukuran *block size* 128 MB, kecepatan rata-ratanya adalah 194.301 ms dibandingkan dengan pada ukuran *block size* 256 MB yakni 188.805 ms dan pada ukuran *block size* 384 MB 158.421 ms (*ISO Image*), sedangkan pada *file Video* dan *Text* dapat dilihat pada Tabel 3.

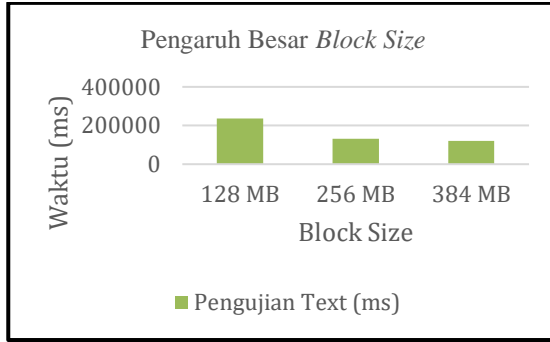


Gambar 10. Pengaruh Besar *Block Size* Terhadap kecepatan pada ukuran *file 2 GB (ISO Image)*

Grafik perbandingan pada Gambar 4.8 dan Gambar 4.9, sehingga dapat disimpulkan bahwa semakin besar *block size* pada *HDFS* maka semakin cepat data didistribusikan. Berikut tampilan kecepatan data dalam bentuk grafik



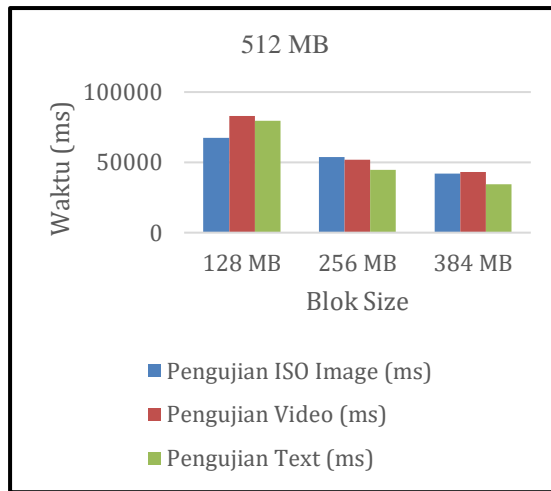
Gambar 11. Pengaruh Besar *Block Size* Terhadap kecepatan pada ukuran *file 2 GB (Video)*



Gambar 12. Pengaruh Besar *Block Size* Terhadap kecepatan pada ukuran *file* 2 GB (*Text*).

3.3 Hasil Pengujian

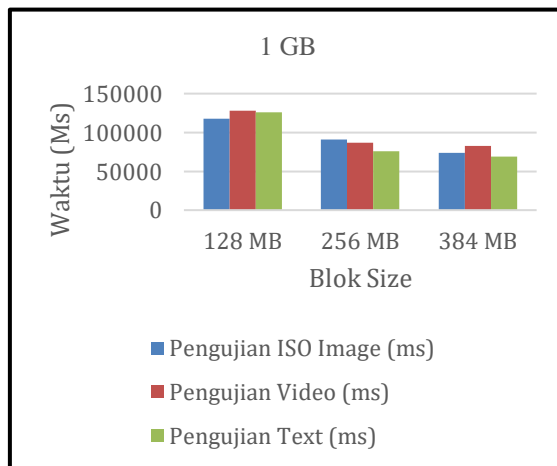
3.3.1 Hasil Pengujian File 512 MB



Gambar 13. Hasil Pengujian File 512 MB

Pada grafik yang tertera pada Gambar 13 dalam pengujian *file* 512 MB tipe data *Text* menjadi tipe data yang memiliki rentetan kecepatan transfer yang lebih baik dari tipe data lainnya.

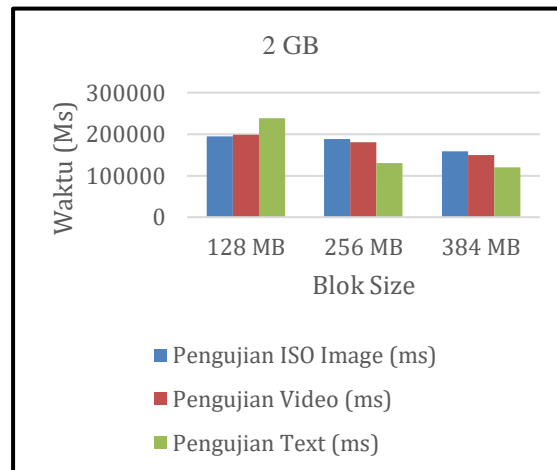
3.3.2 Hasil Pengujian File 1 GB



Gambar 14. Hasil Pengujian File 1 GB

Pada grafik yang tertera pada Gambar 14 dalam pengujian *file* 512 MB tipe data *Text* menjadi tipe data yang memiliki rentetan kecepatan transfer yang lebih baik dari tipe data lainnya.

3.3.3 Hasil Pengujian File 2 GB



Gambar 15. Hasil Pengujian File 2 GB

Pada grafik yang tertera pada Gambar 15 dalam pengujian *file* 512 MB tipe data *Text* menjadi tipe data yang memiliki rentetan kecepatan transfer yang lebih baik dari tipe data lainnya.

4. KESIMPULAN

Berdasarkan hasil penelitian Rancang Bangun *Infrastruktur Big Data* Menggunakan *Hadoop Framework* dapat ditarik kesimpulan.

- 1) Membangun *infrastruktur Big Data* berhasil menggunakan *hadoop framework*.
- 2) *Performance hadoop* pada besar *block size* 384 MB memiliki kecepatan yang lebih baik dari *block size* 128 MB dan 256 MB dikarenakan data akan terbagi menjadi 384 MB sehingga pemetaan data akan lebih sedikit dibanding pada *block size* 128 MB dan 256 MB.
- 3) *Hadoop* sangat baik dalam menangani data bertipe *text* ini dikarenakan data *text* menjadi tipe data yang memiliki rentetan dengan penurunan waktu pengujian yang lebih baik dari tipe data lainnya.

5. REFERENCES

- [1] Apache Hadoop. (<http://hadoop.apache.org/> (online) diakses pada tanggal 20 Agustus 2017).
- [2] Borthakur, Dhruva 2008. *HDFS Architecture Guide*. *Journal Article*.
- [3] Dean, Jeffrey dan Sanjay Ghemawat. 2004. *MapReduce: Simplified Data Processing on Large Clusters*. *Journal Article*.
- [4] Favian, Iqbal Grady. 2017. Analisa Perbandingan *Original Hadoop Cluster* Dan Modifikasi *Hadoop Cluster*. *Journal Article*.
- [5] Grover, Mark, Ted Malaska, Jonathan Seidman, and Gwen Shapira. 2015. *Hadoop Application Architectures*. Sebastopol, United States : O'Reilly Media, Inc.
- [6] Heger, Dominique A. 2013. *Hadoop Design, Architecture & MapReduce Performance*. *Journal Article*.
- [7] Kanellos, Michael. 2016. Amount of Data Created Annually to Reach 180 Zettabytes in 2025, *Article*. (Online). (<https://whatsthebigdata.com/2016/03/07/>, diakses 2 September 2017).
- [8] Khusumanegara, Priagung. 2013. Analisa Pengaruh Block Size Pada Hdfs Terhadap Kecepatan Proses Mapreduce. Seminar. Universitas Indonesia.
- [9] Lam, Chuck. 2011. *Hadoop In Action*. Stamford: Manning Publications Co.
- [10] May, Dima. 2012. *Hadoop Distributed File System (HDFS) Overview*. *Journal Article*.
- [11] Murthy, Arun. 2012. *Apache Hadoop Yarn Concepts and Applications*. *Article (Online)*. (<https://hortonworks.com>, diakses pada tanggal 29 November 2017).
- [12] Novitasari, Rista, Mochamad Hariadi, dan I Ketut Eddy Purnama. 2015. Temu Kembali Informasi Menggunakan Elastic search Pada Unstructured Datatext Multidimensi. *Journal Article*.
- [13] Shvachko, Konstantin, Hairong Kuang, Sanjay Radia, and Robert Chansler. *The Hadoop Distributed File System*. Sunnyvale, California USA, IEEE, 2010.
- [14] Venner, Jason. 2009. *Pro Hadoop*. United States of America : Apress.
- [15] White, Tom. 2015. *Hadoop : The Definitive Guide (4th Edition)*. California: O'Reilly Media, Inc.
- [16] Wjaya, Vijjam. 2013. *HDFS: Berawal dari Google untuk Big Data*. *Article, (Online)*, (<http://www.teknologi-bigdata.com>, diakses pada 23 Agustus 2017).