

**IMPLEMENTASI APACHE SPARK PADA ALGORITMA
BOYER-MOORE HORSPOOL UNTUK STUDI KASUS
*INTERNAL TRANSCRIBED SPACER & RESTRICTION
ENZYME***

SKRIPSI

Diajukan untuk Memenuhi Sebagian dari
Syarat Memperoleh Gelar Sarjana Komputer
Program Studi Ilmu Komputer



Oleh

Fidela Zhafirah

1406424

PROGRAM STUDI ILMU KOMPUTER
DEPARTEMEN PENDIDIKAN ILMU KOMPUTER
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA
BANDUNG
2019

**IMPLEMENTASI APACHE SPARK PADA ALGORITMA
BOYER-MOORE HORSPOOL UNTUK STUDI KASUS *INTERNAL
TRANSCRIBED SPACER & RESTRICTION ENZYME***

Oleh

Fidela Zhafirah

NIM 1406424

Sebuah Skripsi yang Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh
Gelar Sarjana Komputer di Fakultas Pendidikan Matematika dan Ilmu
Pengetahuan Alam

© Fidela Zhafirah 2019

Universitas Pendidikan Indonesia

Juni 2019

Hak Cipta Dilindungi Undang-Undang

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan dicetak
ulang, difoto kopi, atau cara lainnya tanpa izin dari penulis.

FIDELA ZHAFIRAH

1406424

**IMPLEMENTASI SPARK PADA ALGORITMA BOYER-MOORE HORSPOOL
UNTUK STUDI KASUS *INTERNAL TRANSCRIBED SPACER & RESTRICTION
ENZYME***

DISETUJUI DAN DISAHKAN OLEH PEMBIMBING:

Pembimbing I,



Lala Septem Riza, M.T., Ph.D.

NIP. 197809262008121001

Pembimbing II,

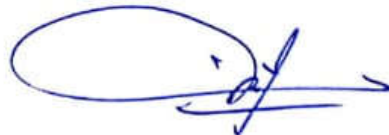


Topik Hidayat, Ph.D.

NIP. 197004101997021001

Mengetahui,

Ketua Departemen Pendidikan Ilmu Komputer



Lala Septem Riza, M.T., Ph.D.

NIP. 197809262008121001

PERNYATAAN

Dengan ini penulis menyatakan bahwa skripsi dengan judul “Implementasi Apache Spark pada Algoritma Boyer-Moore Horspool untuk Studi Kasus *Internal Transcribed Spacer & Restriction Enzyme*” ini beserta seluruh isinya adalah benar-benar karya penulis sendiri. Penulis tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, penulis siap menanggung risiko/sanksi apabila di kemudian hari ditemukan adanya pelanggaran etika keilmuan atau ada klaim dari pihak lain terhadap keaslian karya penulis ini.

Bandung, Juni 2019

Yang Membuat Pernyataan,

Fidela Zhafirah

NIM 1406424

**IMPLEMENTASI APACHE SPARK PADA ALGORITMA
BOYER-MOORE HORSPOOL UNTUK STUDI KASUS *INTERNAL
TRANSCRIBED SPACER & RESTRICTION ENZYME***

Oleh

Fidela Zhafirah — fidela.zhafirah@student.upi.edu

1406424

ABSTRAK

Abstrak— Peningkatan jumlah data yang sangat besar menjadi permasalahan saat ini. Peningkatan jumlah data yang besar membuat penyimpanan sangat besar dan proses data menjadi sangat lama. Sedangkan, kecepatan proses sangat diperlukan untuk mengefisienkan waktu. Penelitian ini ditujukan untuk menyelesaikan permasalahan penyimpanan dan proses sebagai solusi pengolahan data besar dengan membuat model komputasi *string matching* dengan menggunakan algoritma Boyer-Moore Horspool menggunakan Big Data *platform* yaitu Apache Spark dimana Hadoop Distributed File System sebagai penyimpanan data pada *cluster*. Pada penelitian ini dilakukan perbandingan waktu proses pecocokan *string* antara *stand alone*, penggunaan Apache Spark *single node*, penggunaan Apache Spark 3 *nodes*, 5 *nodes*, 11 *nodes* dan 16 *nodes* dengan menggunakan penyimpanan Hadoop Distributed File System pada *cluster* di Google Cloud Platform. Studi kasus yang digunakan adalah bioinformatika dengan menyelesaikan dua permasalahan dalam bidang biologi yaitu pencarian motif terkait dengan penentuan kelompok tanaman berbunga dengan golongan tanaman lainnya dan pencarian motif sebagai deteksi terhadap gejala *begomovirus* sebagai sebab dari penyakit daun keriting. Pada hasil penelitian, diperoleh waktu yang tidak signifikan karena data yang digunakan masih dapat diproses oleh program klasik sehingga waktu eksekusi tidak jauh berbeda. Akurasi dari program yang dijalankan pada Apache Spark ini adalah 83,5%.

Kata Kunci: *Bigdata, Apache Spark, Distributed System, DNA, Bioinformatics*

**APACHE SPARK IMPLEMENTATION IN BOYER-MOORE HORSPOOL
ALGORITHM FOR CASE STUDY OF INTERNAL TRANSCRIBED SPACER
& RESTRICTION ENZYME**

Arranged by

Fidela Zhafirah — fidela.zhafirah@student.upi.edu

1406424

ABSTRACT

Abstract – Increasing the amount of large data makes storage very large and the process of data becomes very long. Meanwhile, the process speed is very necessary to make time efficient. This study aims to solve the problem of storage and process as a large data processing solution by making a string matching computational model using the Boyer-Moore algorithm Horspool using the big data platform, namely Apache Spark where Hadoop Distributed File System as data storage on the cluster. In this study a comparison of the timing of matching strings between stand alone, the use of the Apache Spark single node, the use of Apache Spark 3 nodes, 5 nodes, 11 nodes and 16 nodes by using the Hadoop Distributed File System storage on a cluster on the Google Cloud Platform. The case study used is bioinformatics by solving two problems in the field of biology, namely the search for motifs related to the determination of groups of flowering plants with other plant groups and the search for motifs as detection of symptoms of begomovirus as a cause of curly leaf disease. From the results of the study, the time obtained is not too significant because the data used can still be processed by the classical program so that the execution time is not much different. The accuracy of the program that runs on Apache Spark is 83.5%.

Keywords: Bigdata, Apache Spark, Distributed System, DNA, Bioinformatics

KATA PENGANTAR

Puji serta syukur penulis panjatkan kehadirat Allah SWT., atas kehendak dan izin-Nya-lah penulis dapat menyelesaikan skripsi yang berjudul “Implementasi Apache Spark pada Algoritma Boyer-Moore Horspool untuk Studi Kasus *Internal Transcribed Spacer & Restriction Enzym*” ini tepat pada waktunya. Skripsi ini ditulis untuk memenuhi sebagian syarat dalam meraih gelar sarjana komputer di Program Studi Ilmu Komputer, Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam, Universitas Pendidikan Indonesia.

Selesainya skripsi ini tak lepas dari bantuan banyak pihak. Untuk itu penulis mengucapkan banyak terima kasih untuk pihak-pihak yang telah membantu penulis selama proses pembuatan skripsi.

Akhirnya penulis sampaikan permohonan maaf atas segala kesalahan dalam skripsi ini. Penulis mengharapkan kritik dan saran yang membangun dari pembaca. Semoga skripsi ini bermanfaat bagi kemajuan pendidikan, ilmu pengetahuan dan teknologi. Aamiin.

Bandung, Juni 2019

Penulis

UCAPAN TERIMA KASIH

Alhamdulillahhirabilalamin, puji dan syukur kehadiran Allah SWT. Yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis diberikan kelancaran dalam menyelesaikan penulisan skripsi ini. Dalam proses menyelesaikan penelitian dan penyusunan skripsi ini, peneliti banyak mendapat bimbingan, dorongan, serta bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini peneliti mengucapkan terimakasih serta penghargaan yang setinggi-tingginya, kepada:

1. Kedua orang tua, Papah Farid Abdullah dan Mamah Vishi Aprianti serta adik-adik, Fladio Muhammad Haikal dan Feisya Amelia yang tanpa henti-hentinya memberikan doa dan dukungan, baik itu dukungan moral, materil maupun spiritual sehingga dapat memotivasi penulis dalam menyelesaikan skripsi ini.
2. Bapak Dr. Lala Septem Riza, M.T., selaku pembimbing I atas segala waktu yang dicurahkan untuk membimbing penulis demi terselesaikannya skripsi ini.
3. Bapak Topik Hidayat, Ph.D. selaku pembimbing II yang telah memberikan saran kepada penulis selama proses penyelesaian penelitian dan penulisan skripsi.
4. Bapak Dr. Lala Septem Riza, M.T., selaku Kepala Departemen Pendidikan Ilmu Komputer FPMIPA Universitas Pendidikan Indonesia.
5. Ibu Dr. Rani Megasari, M.T., selaku Ketua Program Studi Ilmu Komputer dan Bapak Jajang Kusnendar, M.T., selaku Ketua Program Studi Pendidikan Ilmu Komputer FPMIPA Universitas Pendidikan Indonesia.
6. Bapak Dr. Wawan Setiawan S.Kom., M.Kom., selaku dosen pembimbing akademik yang telah memberikan arahan dan bimbingan selama penulis menjalani perkuliahan.
7. Bapak dan Ibu Dosen Prodi Pendidikan Ilmu Komputer dan Ilmu Komputer yang telah berbagi ilmu yang sangat bermanfaat kepada penulis.
8. Muhammad Ridwan yang selalu memberikan segala dukungan untuk penulis dalam menjalani hidup, perkuliahan serta proses penulisan skripsi ini.
9. Naufal dan Farhan yang melakukan penelitian di bidang sama telah turut membantu dan saling berbagi penemuan selama proses pengerjaan skripsi.

10. Masa Depan Indah atau Masa Depan Nikah: Faisal Syaiful Anwar, Fikry Al Farisi Muslim, Reinaldy Rahmadian, Muhammad Ridwan, Muhammad Zulfikar, Zakka Muhammad Shiddiq, Agung Cahya Gumilar, Wiwi Juwita, Muhammad Eagan Ramadhan yang selalu mendukung dan memotivasi.
11. Teman seperjuangan Mahasiswa Departemen Pendidikan Ilmu Komputer FPMIPA UPI 2014. Terkhusus untuk teman-teman dari C2 2014 yang telah berjuang dan kenangan indah bersama dari awal masa kuliah hingga selesai.
12. Para asisten Lab Multimedia satu periode: Kang Banyu, Teh Teti, Kang Ferldy, Eagan dan Hisyam.
13. Kang Banyu, Teh Tyas, Kang Arzaaq atas bantuan dan *sharing*-nya dalam terselesaikannya skripsi ini.
14. Neng syva dan team Smile yang selalu menyemangati untuk menyelesaikan skripsi in.
15. Semua pihak yang tidak bisa disebutkan satu persatu yang telah memberi arti dan dukungan pada penulis.

Semoga semua amal baik yang telah diberikan kepada penulis mendapatkan balasan yang berlipat dari Allah SWT. Aamiin

Bandung, Juni 2019

Fidela Zhafirah

DAFTAR ISI

ABSTRAK	i
<i>ABSTRACT</i>	ii
KATA PENGANTAR	iii
UCAPAN TERIMA KASIH.....	iv
DAFTAR ISI.....	vi
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	14
BAB I PENDAHULUAN.....	Error! Bookmark not defined.
1.1. Latar Belakang	Error! Bookmark not defined.
1.2. Rumusan Masalah	Error! Bookmark not defined.
1.3. Tujuan Penelitian.....	Error! Bookmark not defined.
1.4. Manfaat Penelitian.....	Error! Bookmark not defined.
1.5. Batasan Masalah.....	Error! Bookmark not defined.
1.6. Sistematika Penulisan.....	Error! Bookmark not defined.
BAB II KAJIAN PUSTAKA	Error! Bookmark not defined.
2.1. Fenomena Big Data	Error! Bookmark not defined.
2.2. <i>Platform</i> Apache Hadoop.....	Error! Bookmark not defined.
2.2.1. MapReduce	Error! Bookmark not defined.
2.2.2. Hadoop Distributed File System (HDFS)	Error! Bookmark not defined.
2.2.3. Cara Instalasi Apache Hadoop..	Error! Bookmark not defined.
2.2.4. Contoh Program dan Eksekusi pada Apache Hadoop	Error! Bookmark not defined.
2.3. <i>Platform</i> Apache Spark	Error! Bookmark not defined.
2.3.1. Resilient Distributed Dataset (RDD)	Error! Bookmark not defined.
2.3.2. Cara Instalasi Apache Spark	Error! Bookmark not defined.
2.3.3. Contoh Program dan Eksekusi pada Apache Spark.....	Error! Bookmark not defined.
2.4. Bahasa Pemrograman Python	Error! Bookmark not defined.
2.5. Cloud Computing	Error! Bookmark not defined.

2.6.	Google Cloud Platform (GCP)	Error! Bookmark not defined.
2.7.	Penggunaan Teknik <i>String Matching</i>	Error! Bookmark not defined.
2.8.	Algoritma Boyer-Moore	Error! Bookmark not defined.
2.9.	Algoritma Boyer-Moore Horspool	Error! Bookmark not defined.
2.10.	Peneliti Bidang Biologi	Error! Bookmark not defined.
2.11.	Bioinformatika	Error! Bookmark not defined.
2.12.	DNA	Error! Bookmark not defined.
2.13.	<i>Internal Transcribed Spacer (ITS)</i>	Error! Bookmark not defined.
2.14.	Enzim Restriksi	Error! Bookmark not defined.
BAB III METODOLOGI PENELITIAN		Error! Bookmark not defined.
3.1.	Desain Penelitian	Error! Bookmark not defined.
3.2.	Alat dan Bahan Penelitian	Error! Bookmark not defined.
3.2.1.	Alat Penelitian	Error! Bookmark not defined.
3.2.2.	Bahan Penelitian	Error! Bookmark not defined.
3.3.	Metode Penelitian	Error! Bookmark not defined.
3.3.1.	Metode Pengumpulan Data	Error! Bookmark not defined.
3.3.2.	Metode Pengembangan Perangkat Lunak	Error! Bookmark not defined.
BAB IV HASIL DAN PEMBAHASAN		Error! Bookmark not defined.
4.1.	Pengumpulan Data	Error! Bookmark not defined.
4.2.	Perancangan Model Komputasi	Error! Bookmark not defined.
4.2.1.	Perancangan Model Komputasi <i>Stand Alone</i>	Error! Bookmark not defined.
4.2.1.1.	Praproses Data	Error! Bookmark not defined.
4.2.1.2.	Algoritma Boyer-Moore Horspool	Error! Bookmark not defined.
4.2.1.3.	Pos Proses	Error! Bookmark not defined.
4.2.2.	Perancangan Model Komputasi <i>Cluster</i> Apache Spark	Error! Bookmark not defined.
4.2.1.4.	Praproses Data	Error! Bookmark not defined.
4.2.1.5.	Konfigurasi Google Cloud Platform	Error! Bookmark not defined.
4.2.1.6.	Unggah Data ke HDFS	Error! Bookmark not defined.
4.2.1.7.	File di HDFS	Error! Bookmark not defined.
4.2.1.8.	<code>SparkContext ()</code>	Error! Bookmark not defined.
4.2.1.9.	<code>Sc.textFile ()</code>	Error! Bookmark not defined.

4.2.1.10. File di RDD	Error! Bookmark not defined.
4.2.1.11. <code>CreateSkipTable()</code>	Error! Bookmark not defined.
4.2.1.12. Penggunaan <code>Map()</code>	Error! Bookmark not defined.
4.2.1.13. Penggunaan Algoritma Boyer-Moore Horspool	Error! Bookmark not defined.
4.2.1.14. <code>SaveAsTextFile()</code>	Error! Bookmark not defined.
4.3. Pengembangan Perangkat Lunak	Error! Bookmark not defined.
4.3.1. Pengembangan Perangkat Lunak <i>Stand Alone</i>	Error! Bookmark not defined.
4.3.1.1. Analisa	Error! Bookmark not defined.
4.3.1.2. Desain	Error! Bookmark not defined.
4.3.1.3. Implementasi.....	Error! Bookmark not defined.
4.3.1.1. Pengujian Program.....	Error! Bookmark not defined.
4.3.2. Pengembangan Perangkat Lunak <i>Cluster</i> Apache Spark.....	Error! Bookmark not defined.
4.3.2.1. Analisa	Error! Bookmark not defined.
4.3.2.2. Desain	Error! Bookmark not defined.
4.3.2.3. Implementasi.....	Error! Bookmark not defined.
4.3.2.4. Pengujian Program.....	Error! Bookmark not defined.
4.4. Rancangan Skenario Eksperimen	Error! Bookmark not defined.
4.4.1 Skenario Eksperimen Perbandingan Algoritma	Error! Bookmark not defined.
4.4.2 Skenario Eksperimen <i>Stand Alone</i> ...	Error! Bookmark not defined.
4.4.3 Skenario Eksperimen <i>Cluster</i> Apache Spark	Error! Bookmark not defined.
4.5. Hasil Eksperimen	Error! Bookmark not defined.
4.5.1 Hasil Eksperimen Perbandingan Algoritma	Error! Bookmark not defined.
4.5.2 Hasil Eksperimen <i>Stand Alone</i>	Error! Bookmark not defined.
4.5.3 Hasil Eksperimen <i>Cluster</i> Apache Spark	Error! Bookmark not defined.
4.6. Analisa dan Pembahasan Hasil.....	Error! Bookmark not defined.
4.6.1. Perbandingan Kecepatan Algoritma	Error! Bookmark not defined.
4.6.2. Perbandingan Kecepatan Setiap Komputasi	Error! Bookmark not defined.

4.6.3. Perbandingan Nilai Akurasi	Error! Bookmark not defined.
BAB V KESIMPULAN DAN SARAN.....	Error! Bookmark not defined.
5.1. Kesimpulan.....	Error! Bookmark not defined.
5.2. Saran.....	Error! Bookmark not defined.
DAFTAR PUSTAKA	15

DAFTAR TABEL

Tabel 2.1 Mappers' output	Error! Bookmark not defined.
Tabel 2.2 Reducers' input	Error! Bookmark not defined.
Tabel 2.3 Nilai <i>value</i> untuk pergeseran dari karakter <i>pattern</i>	Error! Bookmark not defined.
Tabel 4.1 File sekuens DNA dari FTP NCBI	Error! Bookmark not defined.
Tabel 4.2 Pengujian Program Komputasi <i>Stand Alone</i>	Error! Bookmark not defined.
Tabel 4.3 Pengujian Program <i>Cluster</i> Apache Spark	Error! Bookmark not defined.
Tabel 4.4 Skenario Eksperimen Perbandingan Algoritma...	Error! Bookmark not defined.
Tabel 4.5 Skenario Eksperimen <i>Stand Alone</i>	Error! Bookmark not defined.
Tabel 4.6 Skenario Eksperimen <i>Clustrt</i> Apache Spark.....	Error! Bookmark not defined.
Tabel 4.7 Hasil Eksperimen Perbandingan Algoritma.....	Error! Bookmark not defined.
Tabel 4.8 Hasil Eksperimen <i>Stand Alone</i>	Error! Bookmark not defined.
Tabel 4.9 Hasil Eksperimen <i>Cluster</i> Apache Spark.....	Error! Bookmark not defined.
Tabel 4.10 Nilai Akurasi Hasil Eksperiemn Seluruh Komputasi	Error! Bookmark not defined.

DAFTAR GAMBAR

- Gambar 1.1 Generator Big Data di Internet**Error! Bookmark not defined.**
- Gambar 1.2 Masalah Big Data pada Biologi dan Kedokteran**Error! Bookmark not defined.**
- Gambar 2.1 Atribut-atribut Big Data (Sarraf & Ostadhashem, 2017).....**Error! Bookmark not defined.**
- Gambar 2.2 MapReduce process (Parsian, 2015) ..**Error! Bookmark not defined.**
- Gambar 2.3 HDFS web UI.....**Error! Bookmark not defined.**
- Gambar 2.4 Resources manager web UI.....**Error! Bookmark not defined.**
- Gambar 2.5 Apache Spark *stack* (Karau et al., 2015)**Error! Bookmark not defined.**
- Gambar 2.6 Struktur RDD bekerja di dalam Apache Spark**Error! Bookmark not defined.**
- Gambar 2.7 Penggunaan RDD dalam memuat file teks (Karau et al., 2015) **Error! Bookmark not defined.**
- Gambar 2.8 Hasil *Output* Word Count**Error! Bookmark not defined.**
- Gambar 2.9 Pseudocode Algoritma Boyer-Moore (Zalزالah, 2015) **Error! Bookmark not defined.**
- Gambar 2.10 Peudocode Algoritma Boyer-Moore Horspool (Zalزالah, 2015)**Error! Bookmark not defined.**
- Gambar 2.11 Interaksi disiplin pembentukan bioinformatika (Bayat, 2002) **Error! Bookmark not defined.**
- Gambar 2.12 Identifikasi Molekuler Virus pada CPRE**Error! Bookmark not defined.**
- Gambar 3.1 Desain proses penelitian**Error! Bookmark not defined.**
- Gambar 3.2 Model *Waterfall* (Sommerville, 2011)**Error! Bookmark not defined.**
- Gambar 4.1 Spesies yang tersedia pada laman FTP NCBI**Error! Bookmark not defined.**
- Gambar 4.2 Folder DNA *oryza sativa japonica* pada laman FTP NCBI..... **Error! Bookmark not defined.**

Gambar 4.3 Folder *Assembled_Chromosomes* pada laman FTP NCBI..... **Error! Bookmark not defined.**

Gambar 4.4 Perancangan model komputasi *stand alone***Error! Bookmark not defined.**

Gambar 4.5 Perancangan model *cluster* Apache Spark pada Algoritma Boyer-Moore Horspool**Error! Bookmark not defined.**

Gambar 4.6 *Command prompt* Google Cloud Platform**Error! Bookmark not defined.**

Gambar 4.7 Tampilan UI Hadoop untuk HDFS**Error! Bookmark not defined.**

Gambar 4.8 Lokasi data pada node di HDFS.....**Error! Bookmark not defined.**

Gambar 4.9 Kode Program Header *Stand Alone* ...**Error! Bookmark not defined.**

Gambar 4.10 Kode Program Input Data *Stand Alone***Error! Bookmark not defined.**

Gambar 4.11 Kode Program Fungsi Boyer-Moore Horspool *Stand Alone* ... **Error! Bookmark not defined.**

Gambar 4.12 Pemanggilan Fungsi Boyer-Moore Horspool di Main *Stand Alone***Error! Bookmark not defined.**

Gambar 4.13 Potongan Kode Program Perhitungan dan Penampilan Hasil *Stand Alone***Error! Bookmark not defined.**

Gambar 4.14 Kode Program *Header Cluster* Apache Spark**Error! Bookmark not defined.**

Gambar 4.15 Kode Program Input Data *Cluster* Apache Spark**Error! Bookmark not defined.**

Gambar 4.16 Fungsi `createSkipTable()` *Cluster* Apache Spark..... **Error! Bookmark not defined.**

Gambar 4.17 Fungsi Boyer-Moore Horspool *Cluster* Apache Spark **Error! Bookmark not defined.**

Gambar 4.18 Pemanggilan Fungsi di Main *Cluster* Apache Spark **Error! Bookmark not defined.**

Gambar 4.19 Penghitungan dan Penampilan Hasil *Cluster* Apache Spark ... **Error! Bookmark not defined.**

Gambar 4.20 Perbandingan Kecepatan Waktu antar Algoritma **Error! Bookmark not defined.**

Gambar 4.21 Perbandingan Kecepatan Waktu untuk Skenario A **Error! Bookmark not defined.**

Gambar 4.22 Perbandingan Kecepatan Waktu untuk Skenario B1 **Error! Bookmark not defined.**

Gambar 4.23 Perbandingan Kecepatan Waktu untuk Skenario B2 **Error! Bookmark not defined.**

Gambar 4.24 Perbandingan Kecepatan Waktu untuk Skenario B3 **Error! Bookmark not defined.**

Gambar 4.25 Perbandingan Kecepatan Waktu untuk Skenario B4 **Error! Bookmark not defined.**

Gambar 4.26 Perbandingan Kecepatan Waktu untuk Skenario B5 **Error! Bookmark not defined.**

Gambar 4.27 Perbandingan Kecepatan Waktu untuk Skenario B6 **Error! Bookmark not defined.**

Gambar 4.28 Perbandingan Kecepatan Waktu untuk Skenario C **Error! Bookmark not defined.**

Gambar 4.29 Perbandingan Kecepatan Waktu untuk Skenario D **Error! Bookmark not defined.**

DAFTAR LAMPIRAN

Lampiran 1 Kamus Data Penelitian

Lampiran 2 Cara Penggunaan Google Cloud Platform

DAFTAR PUSTAKA

- Arora, R., & Parashar, A. (2013). Secure User Data in Cloud Computing Using Encryption Algorithms. *International Journal of Engineering Research and Applications (IJERA)*, 3(4), 1922–1926.
- Bayat, A. (2002). Science, medicine, and the future - Bioinformatics. *Clinical Review*, 324(April), 1018–1022.
- Chen, Y., & Wu, Y. (2016). On the massive string matching problem. *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, ICNC-FSKD 2016*, 350–355.
- Dhiba, T. F., Riza, L. S., & Setiawan, W. (2018). Planted Motif Search dalam Sekuens DNA Menggunakan Algoritma Random Projection pada R High-Performance Computing Package.
- Hidayat, T., & Pancoro, A. (2010). Single Strand Conformation Polymorphism Method for Initial Detection DNA Sequences Homogeneity. *HAYATI Journal of Biosciences*, 17(1), 50–52.
- Hidayat, T., Priyandoko, D., Wardiny, P. Y., & Islami, D. K. (2016). Molecular Phylogenetic Screening of *Withania somnifera* Relative From Indonesia Based on Internal Transcribed Spacer Region. *HAYATI Journal of Biosciences*, 23(2), 92–95.
- Ji, C., Li, Y., Qiu, W., Awada, U., & Li, K. (2012). Big Data Processing in Cloud Computing Environments. *Proceedings of the 2012 12th International Symposium on Pervasive Systems, Algorithms, and Networks (i-Span 2012)*, 17–23.
- Karau, H., Konwinski, A., & Wendell, P. (2015). *Learning Spark. Lightning-fast data analysis*.
- Lima, J. W. C., Araújo, R. dos S., Palácio, H. A. de Q., Araújo Neto, J. R. de, & Santos, J. C. N. dos. (2017). Custo Da Erosão Hídrica Para Diferentes Usos Do Solo No Semiárido Brasileiro. *Revista Brasileira de Agricultura Irrigada*, 11(6), 1772–1783.
- Mell, P., & Grance, T. (2011). The NIST Definition of Cloud Computing. *NIST Special Publication 800-145*, 28(12), 1063–1065.
- Phaneendra, S. V., & Reddy, E. M. (2013). Big Data - Solutions for RDBMS Problems - A Survey. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(9), 3686–3691. Retrieved from www.ijarce.com
- Rachman, A. B., Riza, L. S., & Hidayat, T. (2017). Deteksi Genomic Repeats Menggunakan Algoritma Knuth-Morris Pratt pada R High-Performance Computing Package. Bandung: Universitas Pendidikan Indonesia.
- Raita, T. (1994). On tuning the boyer moore horspool string searching algorithm. *Software: Practice and Experience*, 24(4), 435–436.
- S.Vijayarani, & R.Janani. (2017). Information From Desktop – Comparative Analysis. *Bharathiar University*.
- Samsuddin, M., Mat, N., Mazlan, N., Omar, S. N. I., Tajuddin, S., Ngah, N., & Ali, A. M. (2012). Short communication Internal Transcribed Spacer (ITS) as a molecular marker for

- identification of *Dioscorea hispida* Dennst . (Dioscoreaceae). *Journal Agrobiotech*, 3, 47–55.
- Sanner, M. F. (1999). Python: a programming language for software integration and development. *Journal of Molecular Graphics & Modelling*, 17(1), 57–61.
- Sarraf, S., & Ostadhashem, M. (2017). Big data application in functional magnetic resonance imaging using apache spark. *FTC 2016 - Proceedings of Future Technologies Conference*, (December), 281–284.
- Singla, N., & Garg, D. (2012). String Matching Algorithms and their Applicability in various Applications. *Internationall Journal of Soft Computing and Engineering (IJSCE)*, I(6), 218–222.
- Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., ... Robinson, G. E. (2015). Big data: Astronomical or genomical? *PLoS Biology*, 13(7), 1–11.
- van Rossum, G., & Drake, F. L. (2009). *The Python Language Reference Manual*. *Linux J*. (Vol. 1999).
- Vipul, U. (2014). Map-Reduce Parallelization of Motif Discovery.
- White, T. (2015). *Hadoop: The Definitive Guide*. (M. Loukides, M. Blanchette, M. Hacker, & J. Kwityn, Eds.) (4th ed.). California: O'Reilly Media, Inc.
- Wilisiani, F., Somowiyarjo, S., & Hartono, S. (2014). Molecular Identification of Virus Causing Leaf Curl Disease, 18(1), 47–54.
- Wu, X., Zhu, X., Wu, G.-Q., & Ding, W. (2014). Data Mining with Big Data. *IEEE Trasaction On Knowledge And Data Engineering*, 26, 97–107.
- Zalzalalah, C. (2015). *Parallel Boyer-Moore String Matching Algorithm Using Hadoop*. *California State University, Northridge* (Vol. 119). California.