

# Analisis Data Mahasiswa Baru Jurusan Sistem Komputer Dengan Map Reduce

Ahmad Alfarrel Ghazali, Khairunnisya\*, Fauzan Aulia, Ivan Jones Naibaho, Mario Putra Yusandi, Kgs. Muhammad Rizky  
Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya

\*Korespondensi: khairun2210@gmail.com

## ARTICLE INFO

### Article History:

- Received 05 August 2022
- Received in revised form 25 September 2022
- Accepted 19 October 2022
- Available online 30 October 2022

## ABSTRAK

Peran signifikan dari big data dan pembelajaran mesin saat ini menjadi hal yang tak terhindarkan, terutama dalam menganalisis data dan memberikan kecerdasan komputer agar dapat bekerja secara otonom dalam menyelesaikan tugas tertentu. Kemajuan teknologi sensor dan internet telah melimpahkan ketersediaan data, memungkinkan analisis data dalam skala besar. Penelitian ini menggunakan Hadoop dan MapReduce sebagai alat analisis data, memanfaatkan MapReduce sebagai metode untuk mempermudah komputasi pada big data. MapReduce digunakan untuk melakukan komputasi pada kumpulan data yang disimpan dalam Hadoop Distributed File System (HDFS). Metode MapReduce dapat diaplikasikan dalam berbagai bentuk, salah satunya adalah dalam menganalisis data.

Kata Kunci: Big Data, Hadoop, MapReduce, HDFS, Komputasi

## ABSTRACT

*The significant role of big data and machine learning today is inevitable, especially in analyzing data and providing computer intelligence to autonomously perform specific tasks. Advances in sensor technology and the internet have resulted in an abundance of data, enabling large-scale data analysis. This research utilizes Hadoop and MapReduce as tools for data analysis, leveraging MapReduce as a method to facilitate computations on big data. MapReduce is employed to perform computations on datasets stored in the Hadoop Distributed File System (HDFS). The MapReduce method can be applied in various forms, one of which is in data analysis.*

*Keywords: Big Data, Hadoop, MapReduce, HDFS.*

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Big Data merupakan kumpulan data dengan volume yang sangat besar, variasi yang tinggi, pertumbuhan yang cepat, dan mungkin bersifat tidak terstruktur. Proses komputasi pada Big Data dapat melambat jika komputer yang digunakan tidak memenuhi standar yang diperlukan. Oleh karena itu, diperlukan algoritma khusus untuk mendapatkan informasi mendalam dan mendukung pengambilan keputusan yang lebih baik. Sebagai solusi untuk Big Data, Hadoop muncul sebagai sebuah framework open source di bawah lisensi Apache, mendukung aplikasi yang beroperasi pada skala Big Data. Hadoop berasal dari inspirasi makalah Google MapReduce dan Google File System (GFS) yang ditulis oleh ilmuwan Google, Jeffrey Dean dan Sanjay Ghemawat pada tahun 2003. Nama "Hadoop" diberikan oleh Doug Cutting, terinspirasi dari mainan gajah anaknya.

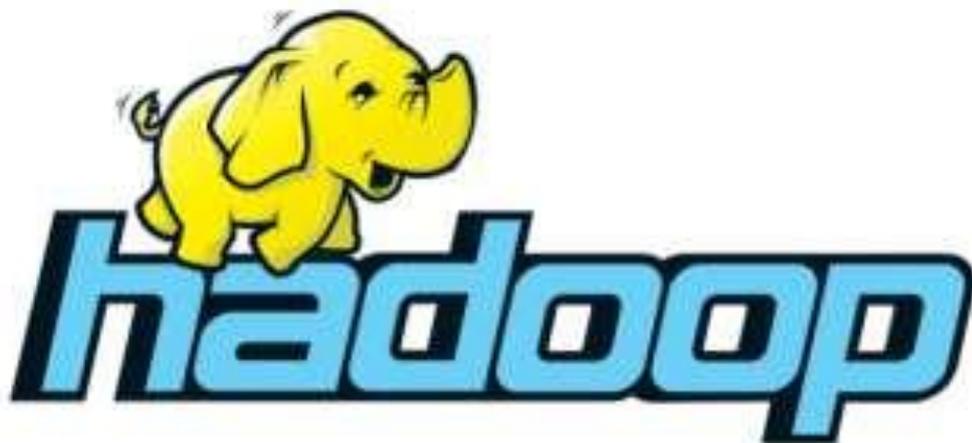
Saat ini, data dihasilkan dengan kecepatan tinggi dari berbagai sumber seperti bisnis, data ilmiah, email, blog, dan sebagainya. Untuk menganalisis dan memproses volume data

besar ini serta mengekstrak informasi yang bermakna, diperlukan aplikasi intensif data dan penyimpanan dalam kluster. Apache Hadoop, sebagai implementasi open source dari sistem MapReduce, menjadi pilihan utama. Hadoop merupakan kerangka kerja perangkat lunak yang dapat diandalkan, terukur, paralel, dan terdistribusi. Sebagai alternatif terhadap sistem mahal, Hadoop memberdayakan pemrosesan paralel pada Big Data dengan menggunakan perangkat keras komoditas. Standard pemrograman Hadoop Distributed File System (HDFS) dan MapReduce cocok untuk pekerjaan analitik intensif Big Data karena arsitektur skala-out dan kemampuannya memproses data secara paralel dalam kluster multi-node. MapReduce, bersama proyek open source Hadoop, memungkinkan pemrosesan paralel data besar dengan partisi otomatis, distribusi data, toleransi kesalahan, dan load balancing[1].

## 2. TINJAUAN PUSTAKA

### 2.1 Hadoop

Hadoop, juga dikenal sebagai Apache Hadoop, merupakan salah satu implementasi teknologi Big Data yang diakses secara open source. Hadoop bukan hanya perangkat lunak biasa, melainkan sebuah kumpulan software yang dapat menangani masalah data dalam jumlah besar. Dengan banyaknya aliran data dalam perkembangan internet saat ini, Hadoop menjadi solusi untuk penyimpanan dan pengelolaan data dalam skala besar. Mengingat volume dan variasi data yang besar yang dikelola oleh perusahaan serta kebutuhan untuk akses data yang cepat, Hadoop diharapkan dapat memberikan solusi yang memadai. Terdapat tiga prinsip utama yang terkait dengan Hadoop. Pertama, Hadoop memiliki kemampuan untuk menggabungkan banyak komputer menjadi satu kesatuan, di mana data akan didistribusikan ke seluruh komputer untuk menjaga keamanan. Kedua, Hadoop memiliki sistem yang memungkinkan pembagian proses perhitungan atau komputasi yang biasanya memakan waktu lama. Secara teknis, Hadoop menggunakan teknik map reduce yang dikendalikan oleh job tracker. Terakhir, sistem Hadoop mampu membagi beban penyimpanan ke berbagai komputer untuk menjaga keamanan data dalam situasi ketika satu komputer mengalami kegagalan. Sistem ini dikenal sebagai Hadoop Distributed File System (HDFS).



Gambar 1. Hadoop

### 2.2. MapReduce

MapReduce adalah kerangka kerja perangkat lunak yang dipatenkan oleh Google untuk mendukung komputasi terdistribusi pada kumpulan data besar di kelompok komputer. Dikenal sebagai mode pemrosesan paralel sederhana, MapReduce memiliki keuntungan signifikan,

seperti kemudahan dalam melakukan komputasi paralel, mendistribusikan data ke prosesor, menjaga keseimbangan di antara mereka, dan menyediakan antarmuka independen dari teknologi backend. MapReduce didesain dengan dua fungsi utama, yaitu Map dan Reduce. Fungsi Map berperan mengambil pasangan input dan menghasilkan set pasangan kunci/nilai perantara. Fungsi Reduce, sementara itu, menerima kunci perantara dan kumpulan nilai untuk kunci tersebut, menggabungkan nilai-nilai ini untuk membentuk kumpulan nilai yang lebih kecil. Dalam prosesnya, fungsi Map mengelompokkan semua nilai antara yang terkait dengan kunci perantara yang sama dan meneruskannya ke fungsi Reduce. Ini menciptakan pengaturan yang efisien untuk memproses dan mengelola kumpulan data yang besar secara terdistribusi.



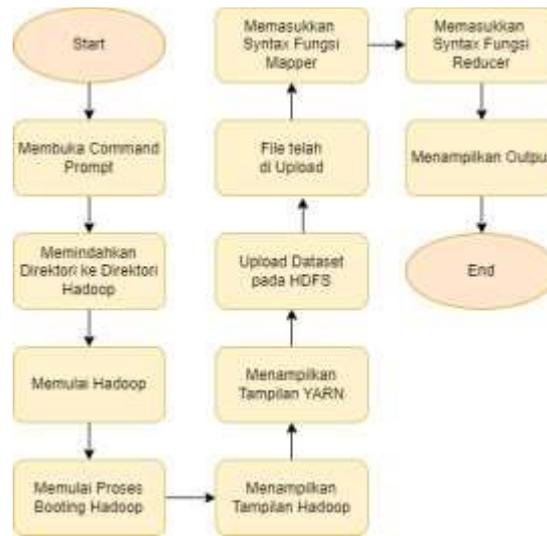
Gambar 2. MapReduce

Kelebihan MapReduce tidak hanya terbatas pada kemampuannya mendistribusikan data ke prosesor dan menjaga keseimbangan, tetapi juga pada antarmuka yang independen dari teknologi backend. Ini menjadikannya alat yang fleksibel dan efektif dalam menyederhanakan dan mempercepat komputasi paralel untuk data besar, memenuhi kebutuhan pemrosesan data yang efisien dan efektif[2].

### 3. METODELOGI PENELITIAN

#### 3.1 Kerangka kerja

Kerangka kerja penelitian merupakan suatu struktur konseptual yang digunakan untuk merancang, mengembangkan, dan menyusun rencana penelitian. Hal ini mencakup langkah-langkah, variabel-variabel, teori-teori, dan metodologi yang akan digunakan dalam penelitian. Kerangka kerja penelitian bertujuan untuk memberikan landasan yang kokoh bagi penelitian dan membimbing peneliti dalam menyusun rencana serta mengarahkan jalannya penelitian tersebut. Dengan merinci unsur-unsur penting, kerangka kerja penelitian membantu peneliti untuk memahami konteks penelitian, menyusun pertanyaan penelitian, dan menentukan metode yang sesuai. Adapun kerangka kerja penelitian ditunjukkan pada gambar berikut ini.



Gambar 3. Kerangka kerja

### 3.2 Jenis dan Pendekatan

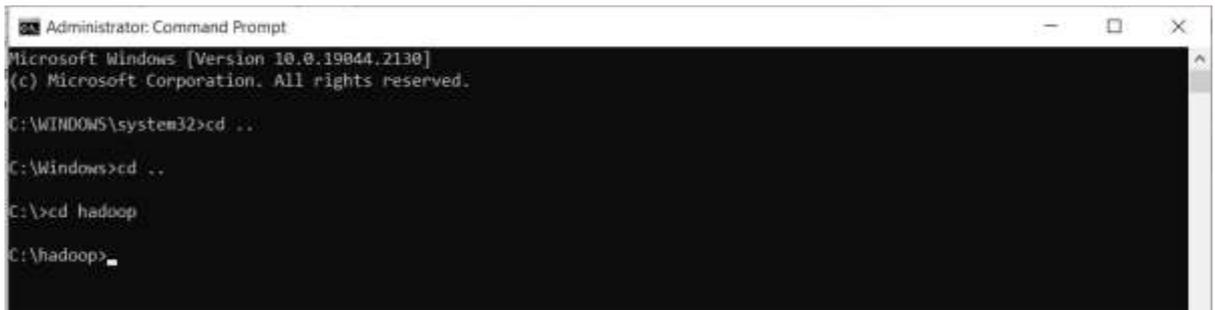
Penelitian ini menggunakan jenis penelitian campuran yang menggabungkan dua pendekatan, yaitu penelitian kualitatif dan penelitian kuantitatif. Kedua jenis penelitian ini diintegrasikan dalam satu kerangka penelitian dengan tujuan memperoleh pemahaman yang lebih komprehensif dan mendetail mengenai masalah yang diteliti. Pendekatan campuran ini melibatkan asumsi filosofis, aplikasi pendekatan kualitatif dan kuantitatif, serta pencampuran data dari keduanya. Integrasi pendekatan kualitatif dan kuantitatif dalam satu penelitian menimbulkan kompleksitas analisis, karena melibatkan pengumpulan dan analisis dua jenis data yang berbeda.

Penelitian ini fokus pada analisis data mahasiswa baru jurusan sistem komputer menggunakan pendekatan campuran. Pendekatan campuran merupakan gabungan antara penelitian kuantitatif dan kualitatif, yang menjadikannya lebih kompleks daripada penelitian yang hanya menggunakan satu jenis pendekatan. Dalam penelitian ini, tidak hanya mengumpulkan dan menganalisis data, tetapi juga melibatkan fungsi-fungsi kuantitatif dan kualitatif. Penggunaan kedua metode ini bertujuan untuk memberikan pemahaman yang lebih holistik terkait dengan masalah penelitian. Pendekatan campuran memiliki tujuan untuk memahami isu dengan mengartikulasikan data kualitatif sebagai rincian deskriptif dan data kuantitatif sebagai angka-angka. Metode penelitian campuran juga bertujuan untuk mendapatkan hasil statistik kuantitatif dari objek penelitian tertentu, yang kemudian diikuti dengan pengamatan individu untuk mendapatkan pemahaman yang lebih mendalam.

### 3.3 Langkah-langkah Percobaan

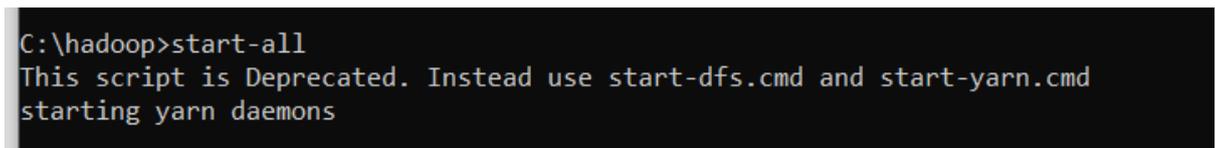
Pada percobaan ini diperlukan JDK versi 8 untuk menginstall Hadoop pada windows. Setelah Hadoop terinstall dan terkonfigurasi pada komputer maka percobaan dapat dilakukan.

1. Buka Command Promt dengan run as administrator, setelah itu berpindah ke direktori yang terdapat Hadoop.



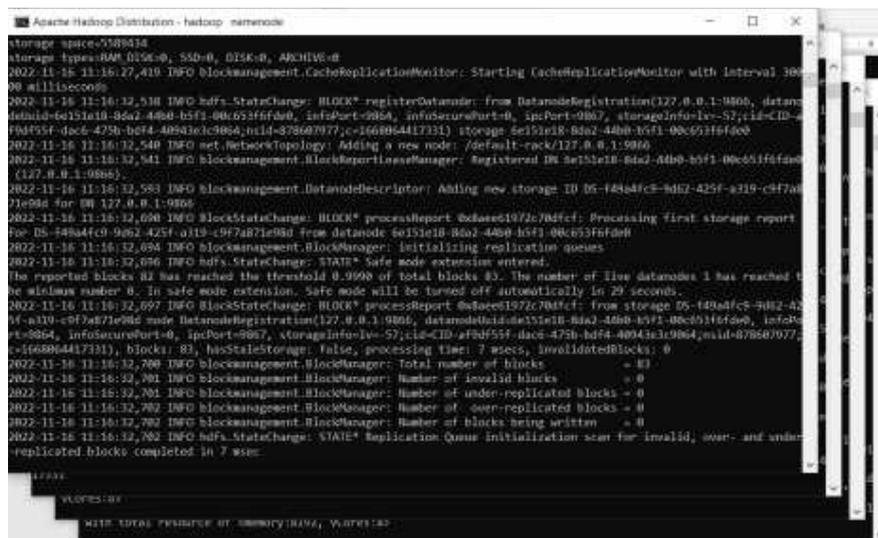
Gambar 4. Terminal Hadoop

2. Memulai Hadoop



Gambar 5. Menjalankan Hadoop

Setelah di-enter maka akan terbuka 4 pop-up command prompt seperti gambar di bawah ini, Jika 4 command prompt tersebut tidak ada yang berhenti/stop, maka dapat disimpulkan bahwa Hadoop telah sukses dijalankan.



Gambar 6. Proses booting Hadoop

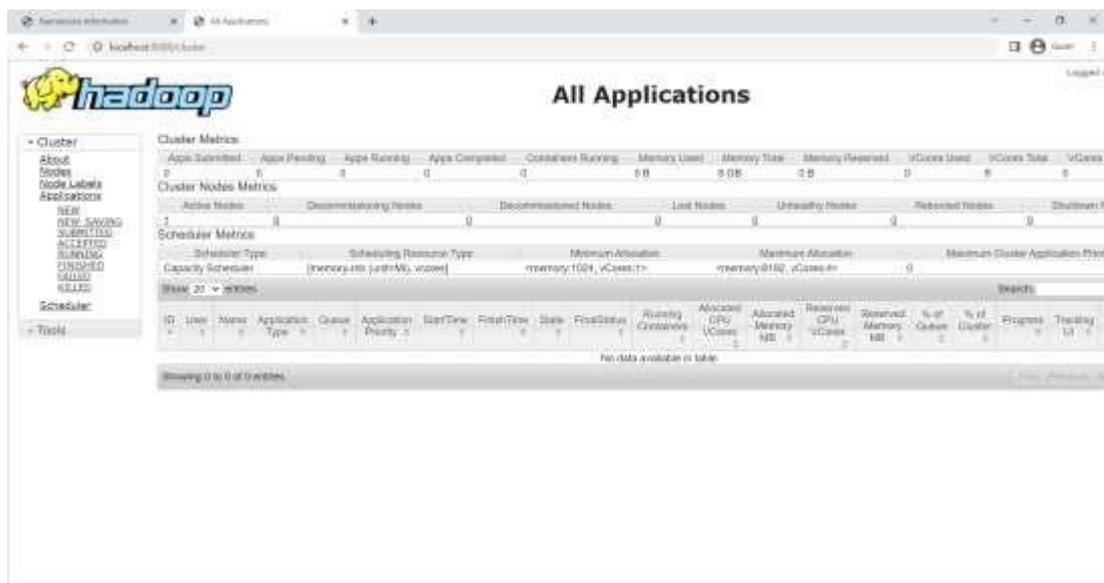
3. Mengakses UI Hadoop pada Browser

Setelah Hadoop berhasil/sukses dijalankan pada command prompt. Selanjutnya untuk mengakses Hadoop tersebut, perlu menggunakan sebuah browser. Berikut cara pengaksesan Hadoop menggunakan browser

- a. Hadoop NameNode UI : <https://localhost:9870>
- b. YARN ResourceManager : <https://localhost:8088>



Gambar 7. Tampilan awal Hadoop



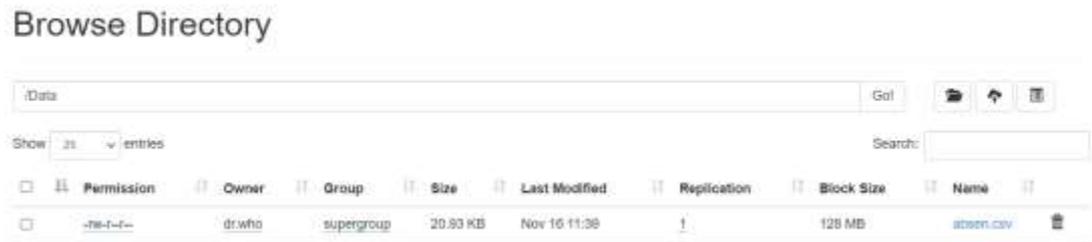
Gambar 8. Tampilan YARN

4. Memasukkan Data pada HDFS  
Data dapat dimasukkan ke dalam Hadoop Distributed File System (HDFS) menggunakan command prompt atau antarmuka pengguna Hadoop, seperti yang terlihat pada Gambar di bawah ini.



Gambar 9. Upload dataset pada HDFS

5. Setelah data terunggah, dapat dilihat daftar file yang telah di-upload seperti pada Gambar berikut.



Gambar 10. Menampilkan file yang telah ter-upload

## 6. Memproses Data

Pemrosesan data memerlukan kode atau program untuk Mapper dan Reducer. Dalam hal ini, digunakan bahasa pemrograman Python dengan file bernama Mapper.py dan Reduce.py untuk melakukan pemrosesan data. Syntax fungsi Mapper dan Reduce dapat dilihat pada Gambar 13 di bawah ini.

### Mapper.py

```
1 #!/usr/bin/env python
2
3 import sys
4
5
6 for line in sys.stdin:
7     drop(axis=0)
8     line = line.strip()
9     words = line.split(",")
10    try:
11        print ('%s|%s' % (words[4],words[2].strip()))
12    except:
13        pass
```

Gambar 11. Syntax fungsi Mapper

### Reduce.py

```
1 #usr/bin/env python
2
3 from operator import itemgetter
4 import sys
5 from operator import itemgetter
6
7 current_word = None
8 Current_good = ""
9 words = None
10
11 for line in sys.stdin:
12     line = line.strip()
13     words, good = line.split('|')
14     if current_word == words:
15         current_good = current_good + "," + good
16     else:
17         if current_word:
18             print('%s|%s' % (current_word, current_good))
19             current_good = good
20             current_word = words
21
22 if current_word == words:
23     print('%s|%s' % (current_word,current_good))
24
```

Gambar 12. Syntax fungsi Reduce

7. Untuk menjalankan proses Map Reduce, diperlukan perintah seperti yang terlihat pada Gambar berikut.

```
C:\hadoop>hadoop jar share/hadoop/tools/lib/hadoop-streaming-*.jar -files "file:///C:/hadoop/Mapper.py,file:///C:/hadoop/Reduce.py" -mapper "python Mapper.py" -reducer "python Reduce.py" -input /Data/absen.csv -output /hasil/045
```

Gambar 13. Command untuk menjalankan proses Map Reduce

- Setelah proses selesai, hasil data dapat diunduh, dan lokasinya dapat dilihat pada Gambar di bawah ini.



Gambar 14. Download hasil data yang telah diproses

"Cepusakti",	"Rana2@gmail",	"Sesatengkap",	"WDM",	"Donjallikepu",	"JalurMasuk",	"Angkatan"
"2022/08/11 10:43:30 AM GMT+7",	"hammad.schl@gmail.com",	M. Hamis Al Bughis,	"09011382227162",	"Bukit",	"Mandiri",	"2022"
"2022/08/11 10:43:49 AM GMT+7",	"rickyriwaido77@gmail.com",	Ricky Rivaldo,	"09011382227036",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:43:54 AM GMT+7",	"dicapriosanvic@gmail.com",	Sanvic Dicaprio,	"09011382227081",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:43:57 AM GMT+7",	"fuadalmadile@gmail.com",	Fuad Almadila Muhammad,	"09011382227030",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:44:00 AM GMT+7",	"fatburfarisanyah@gmail.com",	Muhammad Febhor Farisanyah,	"09011382227141",	"Bukit",	"Mandiri",	"2022"
"2022/08/11 10:44:06 AM GMT+7",	"Amaliansarub@gmail.com",	"Amalia Saruba",	"09011382227059",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:44:06 AM GMT+7",	"aryadikainron@gmail.com",	"Aryadika",	"09011382227032",	"Bukit",	"Mandiri",	"2022"
"2022/08/11 10:44:06 AM GMT+7",	"nabiliasuci2@gmail.com",	"Nabilla Suci Febriani",	"09011382227019",	"Indralaya",	"SBMPTN",	"2022/2023"
"2022/08/11 10:44:06 AM GMT+7",	"ricorikodireal11111@gmail.com",	"M. Rio Erebana",	"09011382227121",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:44:10 AM GMT+7",	"shawwarigawin@gmail.com",	"M. Gavin Hawwar",	"09011382227035",	"Indralaya",	"SBMPTN",	"22"
"2022/08/11 10:44:14 AM GMT+7",	"tsamiatuzachab@gmail.com",	"Tsamiatu Zahrah Azizah",	"09011382227033",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:44:20 AM GMT+7",	"shandycox@gmail.com",	"SHANDY",	"09011382227130",	"Bukit",	"Mandiri",	"2022"
"2022/08/11 10:44:20 AM GMT+7",	"revidyaa@gmail.com",	"Revidya Aprilia Sandiva",	"09011382227054",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:44:25 AM GMT+7",	"rizingputrasari306@gmail.com",	"Ririn Purnama Sari",	"09011382227046",	"Indralaya",	"SBMPTN",	"2021"
"2022/08/11 10:44:30 AM GMT+7",	"alfathulakbar12@gmail.com",	"Muhammad Alfathul Akbar",	"09011382227153",	"Bukit",	"Mandiri",	"2022"
"2022/08/11 10:44:34 AM GMT+7",	"casyqzharfira@gmail.com",	"Casyqa Zharfira",	"09011382227031",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:44:37 AM GMT+7",	"kcpurriapriia04@gmail.com",	"Egga Achida Kirana",	"09011382227168",	"Bukit",	"Mandiri",	"2022"
"2022/08/11 10:44:41 AM GMT+7",	"sigitjambi18@gmail.com",	"SIGIT SULESTYO",	"09011382227027",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:44:43 AM GMT+7",	"allifyal3@gmail.com",	"Allifia Gothrunnada",	"09011382227040",	"Indralaya",	"SBMPTN",	"2022/2023"
"2022/08/11 10:44:46 AM GMT+7",	"achmadzakisyawallianto09@gmail.com",	"Achmad Seki Syawallianto",	"09011382227151",	"Bukit",	"Mandiri",	"2022"
"2022/08/11 10:44:49 AM GMT+7",	"albhaghow199@gmail.com",	"Zhu Ribben Albhaghow",	"09011382227024",	"Indralaya",	"SBMPTN",	"2022/2023"
"2022/08/11 10:44:51 AM GMT+7",	"kopurriapriia04@gmail.com",	"Egga Achida Kirana",	"09011382227020",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:44:54 AM GMT+7",	"bima.arief2016@gmail.com",	"Bima Arief Nugraha",	"09011382227157",	"Bukit",	"Mandiri",	"2023"
"2022/08/11 10:44:57 AM GMT+7",	"akbarsaputra1@gmail.com",	"M. AKSAR SAPUTRA",	"09011382227088",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:44:58 AM GMT+7",	"keysafikrinuharse@gmail.com",	"Keyza Fikri Mubarsa",	"09011382227043",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:45:02 AM GMT+7",	"Mirasajal96@gmail.com",	"Mirinda Gomasul Jaya",	"09011382227110",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:45:07 AM GMT+7",	"gulsandanyas2@gmail.com",	"Maya Yolanda",	"09011382227005",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:45:12 AM GMT+7",	"iszotihabila944@gmail.com",	"Iszoti Habila",	"09011382227079",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:45:13 AM GMT+7",	"sheriniasuryandari@gmail.com",	"Sherina suryandari",	"09011382227016",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:45:17 AM GMT+7",	"sherynelianwellea31114@gmail.com",	"SHERLI AMELIA",	"09011382227097",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:45:17 AM GMT+7",	"ast113823@gmail.com",	"Intan Permatahati",	"09011382227036",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:45:23 AM GMT+7",	"salfitiasa13@gmail.com",	"Sagitri",	"09011382227138",	"Bukit",	"Mandiri",	"2022"
"2022/08/11 10:45:24 AM GMT+7",	"shetudinia@gmail.com",	"Shetudinia",	"09011382227058",	"Indralaya",	"SBMPTN",	"2022/2023"
"2022/08/11 10:45:29 AM GMT+7",	"meydarsa101@gmail.com",	"Riki Meida Sari",	"09011382227067",	"Indralaya",	"SBMPTN",	"2022/2023"
"2022/08/11 10:45:32 AM GMT+7",	"adnanrifi49@gmail.com",	"M. Adnan Rifki Agustian",	"09011382227100",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:45:33 AM GMT+7",	"dezkyputerasatrio@gmail.com",	"Dezky Putera Satrio",	"09011382227042",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:45:37 AM GMT+7",	"karsinawiy2019@gmail.com",	"Prince Karins Wijaya",	"09011382227051",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:45:38 AM GMT+7",	"septianwidia@gmail.com",	"Septiani Widia Rosalia",	"09011382227047",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:45:39 AM GMT+7",	"Alansyah393@gmail.com",	"Alansyah Putra",	"09011382227114",	"Indralaya",	"SBMPTN",	"2022-2023"
"2022/08/11 10:45:40 AM GMT+7",	"zihnithallah54@gmail.com",	"Muhammad Zihni Athallah",	"09011382227084",	"Indralaya",	"SBMPTN",	"2022"
"2022/08/11 10:45:41 AM GMT+7",	"dunamyan12@gmail.com",	"Dunam Febrlan Pratama",	"09011382227167",	"Bukit",	"Mandiri",	"2022"
"2022/08/11 10:45:44 AM GMT+7",	"alyssakhairun@gmail.com",	"Alyssa Khairunisa Maharani",	"09011382227118",	"Indralaya",	"SBMPTN",	"22"
"2022/08/11 10:45:50 AM GMT+7",	"zdmnp1@gmail.com",	"Muhammad Alzidan Pratama",	"09011382227057",	"Indralaya",	"SBMPTN",	"2022"

Gambar 15. Data asli/awal

Tampilan dari gambar di atas adalah data asli pada percobaan yang dilakukan. Untuk melihat data hasil yang telah diproses, dapat dilihat pada gambar berikut.

"Bukit",	"Achmad Seki Syawallianto",	"Ahmad Husayni Denuarta",	"Ahmad Husayni Denuarta",	"Annisa Ulul Azmi Dewi",	"Aryadika",	"BAJU TRISAPUTRO",	"Bima Arief Nugraha",	"Chika Putri Fajarini",	"Donjallikepu",	"Hesatengkap",
"Indralaya",	"ANDREA UCI LENIYA",	"AKHAR FAH REKI",	"Ahmad Azri Jundi",	"Alansyah Putra",	"Aldi Dudira",	"Allifia Gothrunnada",	"Almifah Sellysta Aurelle",	"Alyssa Khairunisa Maharani",	"B",	"R

Gambar 16. Hasil MapReduce

#### 4. HASIL PENELITIAN

Hasil penelitian menunjukkan bahwa data yang telah diproses menggunakan program Mapper dan Reducer menghasilkan output yang merepresentasikan penggabungan data dari input melalui command. Semua data yang telah diolah menampilkan hasil yang terstruktur dan relevan. Proses pengolahan dapat dilakukan melalui command prompt atau antarmuka pengguna Hadoop.

Proses memasukkan data ke dalam Hadoop Distributed File System (HDFS) dapat dilakukan menggunakan command prompt atau antarmuka pengguna Hadoop. Setelah data diunggah, tahap selanjutnya melibatkan pemrosesan data dengan menggunakan program Mapper dan Reducer yang ditulis dalam bahasa pemrograman Python, dengan file bernama Mapper.py dan Reduce.py. Syntax contoh dari fungsi Mapper dapat dilihat pada Gambar 13, sedangkan untuk fungsi Reducer, syntax contohnya ditunjukkan pada Gambar 14.

Setelah proses pemrosesan selesai, perlu dilakukan perintah tertentu untuk menyimpan output. Perintah ini akan menyimpan hasil pengolahan dalam direktori yang ditentukan. Selanjutnya, hasil data yang telah diproses dapat diunduh untuk dianalisis lebih lanjut. Proses ini memungkinkan pemahaman yang lebih baik terhadap masalah penelitian yang diajukan.

#### 5. KESIMPULAN

Penelitian ini menyimpulkan beberapa hal penting terkait proses komputasi pada big data dan pemanfaatan Hadoop sebagai solusi untuk menangani data dalam skala besar. Ditemukan bahwa lambatnya proses komputasi pada big data dapat disebabkan oleh kurangnya standar yang dipenuhi oleh komputer pemroses data. Hadoop, sebagai kumpulan software, memberikan solusi efektif untuk menyelesaikan permasalahan yang muncul dalam pemrosesan data berskala besar. Salah satu keunggulan Hadoop terletak pada kemampuannya untuk mengintegrasikan banyak komputer menjadi satu kesatuan. Penggabungan ini memungkinkan data tersebar di seluruh komputer, menciptakan sistem yang tangguh dan aman. Hadoop juga berperan sebagai kerangka kerja yang berfungsi sebagai tempat penyimpanan dan pemrosesan dataset. Modul Hadoop MapReduce, salah satu komponennya, mampu membagi pekerjaan menjadi tugas-tugas yang dapat dilakukan secara paralel. Pada penelitian ini, diterapkan teknik pemisahan data berdasarkan domisili kampus mahasiswa baru jurusan sistem komputer menggunakan pendekatan MapReduce.

#### DAFTAR PUSTAKA

- [1]. Ghazi, M.R. and Gangodkar, D. (2015) 'Hadoop, mapreduce and HDFS: A developers perspective', *Procedia Computer Science*, 48(C), pp. 45–50. [Online]. Available at: <https://doi.org/10.1016/j.procs.2015.04.108>.
- [2]. Ilmu, U. and Kunci-asosiasi, K. (no date) 'MapReduce sebagai Model Pemrograman untuk Algoritma Aturan Asosiasi pada Hadoop'.
- [3]. Nelfira, N. (2017) 'Rancang Bangun Aplikasi Pembelajaran Sistem Operasi Windows Pada Matakuliah Sistem Operasi Di STMIK Indonesia Padang Berbasis Multimedia Interaktif', *Edik Informatika*, 2(2), pp. 182–189. [Online]. Available at: <https://doi.org/10.22202/ei.2016.v2i2.1461>.

- [4]. Taqwin, P.A.T., Osmond, A.B. and Latuconsina, R. (2018) ‘Implementasi Metode Mapreduce Pada Big Data Berbasis Hadoop Distributed File System’, e-Proceeding of Engineering, 5(1), pp. 1013–1020. [Online]. Available at: <https://doi.org/10.1109/ACDT.2017.7886152>.
- [5]. V. Diaconita, A. R. Bologna, and R. Bologna, “Hadoop oriented smart cities architecture,” Sensors (Switzerland), 2018, doi: 10.3390/s18041181.
- [6]. Rudianto, Arief. (2011). Pemrograman Web Dinamis menggunakan PHP dan MySQL. C.V ANDI OFFSET. Yogyakarta.