

TUGAS AKHIR - KM4801

PENGEMBANGAN PROTOTIPE BIG DATA PADA EKOSISTEM HADOOP MENGGUNAKAN HDFS DAN MAPREDUCE SEBAGAI MODEL KOMPUTASI PARALEL (STUDI KASUS: SAMPEL SISTEM TRANSPORTASI KOTA SURABAYA)

M. FIAN FACHRY A. NRP 06111440000068

Dosen Pembimbing Dr. Budi Setiyono, S.Si, MT

DEPARTEMEN MATEMATIKA Fakultas Matematika, Komputasi, dan Sains Data Institut Teknologi Sepuluh Nopember Surabaya 2019



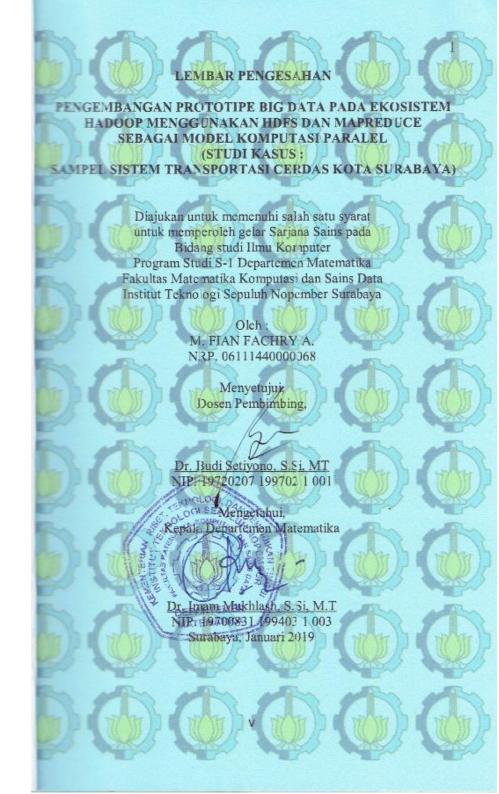
FINAL PROJECT - KM4801

BIG DATA PROTOTYPE DEVELOPMENT ON HADOOP ECOSYSTEM USING HDFS AND MAPREDUCE AS THE PARALLEL COMPUTATION MODEL (CASE STUDY: SAMPLE OF INTELLIGENT TRANSPORTATION SYSTEM OF SURABAYA CITY)

M. FIAN FACHRY A. NRP 06111440000068

Supervisor Dr. Budi Setiyono, S.Si, MT

DEPARTMENT OF MATHEMATICS Faculty of Mathematics, Computation, and Data Science Institut Teknologi Sepuluh Nopember Surabaya 2019





PENGEMBANGAN PROTOTIPE BIG DATA PADA EKOSISTEM HADOOP MENGGUNAKAN HDFS DAN MAPREDUCE SEBAGAI MODEL KOMPUTASI PARALEL (STUDI KASUS : SAMPEL SISTEM TRANSPORTASI KOTA SURABAYA)

Nama Mahasiswa : M. Fian Fachry A. NRP : 06111440000068

Departemen : Matematika

Dosen Pembimbing : Dr. Budi Setiyono, S.Si, MT

Abstrak

Dengan perkembangan sosial yang begitu cepat, industri transportasi juga menghadapi tantangan yang belum pernah dihadapi sebelumnya. Salah satu daerah yang mengalami dampak tersebut adalah Kota Surabaya, yaitu tantangan pertumbuhan jumlah kendaraan pribadi yang begitu pesat. Maka dari itu dibutuhkan suatu sistem transportasi umum yang memadai, yaitu dengan menyediakan rute yang sesuai dengan kebutuhan masyarakat Kota Surabaya. Dengan mengamati jalur perjalanan kendaraan yang berlalu – lalang di Kota Surabaya, akan diketahui preferensi jalur perjalanan yang dilalui oleh masyarakat, sehingga dapat dijadikan acuan untuk pemerintah Kota Surabaya, khususnya Dinas Perhubungan, sebagai dasar membuka rute transportasi umum dalam Kota Surabaya. Dengan memanfaatkan big data, maka data jalur perjalanan kendaraan dapat diolah sehingga mendapatkan hasil jalur perjalanan kendaraan yang paling sering dilalui oleh masyarakat Kota Surabaya. Big data tersebut diterapkan pada ekosistem Hadoop dengan menggunakan HDFS dan MapReduce sebagai model komputasi paralel. Model komputasi paralel digunakan karena data yang diolah berjumlah besar, sehingga akan lebih efisien dari model komputasi sekuensial. Dikarenakan area Kota Surabaya yang begitu luas dan prasarana untuk mengambil data yang dibutuhkan kurang memadai, sehingga penelitian ini dilakukan pada sampel 5 titik CCTV, yaitu Jl. Kayoon, Jl. Panglima Sudirman, Jl. Urip Sumoharjo, Keputran, dan Jl. Raya Ngagel. Berdasarkan hasil pengolahan data pada 5 titik

CCTV tersebut, didapatkan bahwa jalur perjalanan kendaraan yang paling sering dilalui adalah jalur perjalanan dari Jl. Panglima Sudirman menuju Jl. Urip Sumoharjo.

Kata kunci : big data, sistem transportasi, implementasi, teknologi, transportasi umum, hadoop, HDFS, MapReduce

BIG DATA PROTOTYPE DEVELOPMENT ON HADOOP ECOSYSTEM USING HDFS AND MAPREDUCE AS THE PARALLEL COMPUTATION MODEL (CASE STUDY: SAMPLE OF INTELLIGENT TRANSPORTATION SYSTEM OF SURABAYA CITY)

Name of Student : M. Fian Fachry A. NRP : 06111440000068

Department : Mathematics

Supervisor : Dr. Budi Setiyono, S.Si, MT

Abstract

Because of the rapid social development, the transportation industry also faces challenges that have never been faced before. One of the region that experienced this impact is the Surabaya City, that is the challenge of the rapid growth of the number of private vehicles. Therefore, an adequate public transportation system is needed, that is by providing a route that fits the needs of the society of Surabaya City. By observing the travel path of passing vehicles in the Surabaya City, the route path preferences traveled by the society will be known, so that it can be used as a reference for the Surabaya City government, especially the Department of Transportation, as the basis for opening public transportation routes in Surabaya City. By utilizing big data, the vehicle path data can be processed so that the results of the vehicle paths that are most frequently traveled by the society of Surabaya City. Big data is applied to the Hadoop ecosystem using HDFS and MapReduce as the parallel computation model. The parallel computation model is used because the processed data is large, so it will be more efficient than the sequential computing model. Due to the vast area of the Surabaya City and the infrastructure to acquire the data needed is not sufficient, so this research was conducted on samples of 5 CCTVs points, those are Kayoon street, Panglima Sudirman street, Urip Sumoharjo street, Keputran, and Raya Ngagel street. Based on the results of data processed at the 5 CCTVs points, it was found that the

vehicle's most frequently traveled route was the travel route from Panglima Sudirman street headed for Urip Sumoharjo street.

Key words: big data, transportation system, implementation, technology, public transportation, hadoop, HDFS, MapReduce

KATA PENGANTAR

Segala puji syukur penulis panjatkan ke hadirat Allah SWT, karena dengan ridho-Nya penulis dapat menyelesaikan Tugas Akhir yang berjudul

"PENGEMBANGAN PROTOTIPE BIG DATA PADA EKOSISTEM HADOOP MENGGUNAKAN HDFS DAN MAPREDUCE SEBAGAI MODEL KOMPUTASI PARALEL (STUDI KASUS: SAMPEL SISTEM TRANSPORTASI CERDAS KOTA SURABAYA)"

Merupakan salah satu persyaratan akademis dalam menyelesaikan Program Sarjana Departemen Matematika, Fakultas Matematika Komputasi dan Sains Data, Institut Teknologi Sepuluh Nopember Surabaya.

Tugas Akhir ini dapat diselesaikan dengan baik berkat kerja sama, bantuan, dan dukungan dari banyak pihak. Sehubungan dengan hal tersebut, penulis ingin mengucapkan terimakasih kepada:

- 1. Dr. Imam Mukhlas, S.Si, MT selaku Kepala Departemen Matematika ITS.
- 2. Dr. Dra. Mardlijah, MT selaku Dosen Wali yang telah memberikan arahan akademik selama penulis menempuh pendidikan di Departemen Matematika ITS.
- 3. Dr. Budi Setiyono, S.Si, MT selaku Dosen Pembimbing yang telah memberikan bimbingan dan motivasi kepada penulis dalam mengerjakan Tugas Akhir ini sehingga dapat terselesaikan dengan baik.
- 4. Dr. Didik Khusnul Arif, S.Si, M.Si selaku ketua program studi S1 Departemen Matematika ITS.
- 5. Drs. Iis Herisman, M.Si selaku Sekretaris Program Studi S1 Departemen Matematika ITS.
- 6. Seluruh jajaran dosen dan staf Departemen Matematika ITS.
- 7. Orang tua dan kedua adik penulis yang senantiasa memberikan dukungan dan do'a yang tak terhingga.

8. Semua pihak yang tak bisa penulis sebutkan satu-persatu, terima kasih telah membantu sampai terselesaikannya Tugas Akhir ini.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan kritik dan saran dari pembaca. Akhir kata, semoga Tugas Akhir ini dapat bermanfaat bagi semua pihak yang berkepentingan.

Surabaya, Januari 2019

Penulis

DAFTAR ISI

KATA PENGANTAR	xi
DAFTAR ISI	. xiii
DAFTAR GAMBAR	xv
DAFTAR TABEL	. xvi
BAB I. PENDAHULUAN	1
1.1 Latat Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan	3
1.5 Manfaat	
1.6 Sistematikan Penulisan Tugas Akhir	3
BAB II. TINJAUAN PUSTAKA	5
2.1. Big Data	5
2.2. Hadoop	7
2.2.1. MapReduce	9
2.2.2. HDFS	
2.3. Implementasi Big Data pada ITS	
2.4. Kondisi Sistem Lalu Lintas Cerdas Kota Surabaya	
BAB III. METODE PENELITIAN	
3.1 Tahapan Penelitian	
3.2 Objek dan Aspek Penelitian	
3.3 Blok Diagram	
BAB IV. PERANCANGAN DAN IMPLEMENTASI SISTEM	
4.1 Perancangan Sistem	
4.1.1. Perancangan Data	
4.1.2. Perancangan Blok Diagram Sistem	
4.1.2.1. Akuisisi Data (Pra-Proses)	
4.1.2.2. Arsitektur Sistem	
4.1.2.3. Perancangan Model Paralel	28
4.1.2.4. Perancangan Struktur Data dan	
Penyimpanan Data pada HDFS	
4.1.2.5. Perancangan Fungsi Mapper	
4.1.2.6. Perancangan Fungsi Reducer	
4.1.3. Perancangan <i>Class</i> Diagram	39

4.2 Implementasi Sistem	40
4.2.1. Implementasi Struktur Data dan Penyimpanan	
Data pada HDFS	40
4.2.2. Implementasi Fungsi Mapper	41
4.2.3. Implementasi Fungsi Reducer	45
4.2.4. Running Program	45
BAB V. HASIL UJI COBA DAN PEMBAHASAN	49
5.1 Hasil Akuisisi Data	49
5.2 Hasil Uji Coba Sistem	49
5.3 Analisa Hasil Pengolahan Data	50
BAB VI. PENUTUP	
6.1 Kesimpulan	53
6.2 Saran	53
DAFTAR PUSTAKA	55
LAMPIRAN	57

DAFTAR GAMBAR

Gambar 2.1.	Pertumbuhan dan digitalisasi kapasitas	
	penyimpanan informasi global	6
Gambar 2.2.	Sebuah cluster multi-node Hadoop	
Gambar 2.3.	Contoh MapReduce kanonik menghitung	
	kemunculan setiap kata dalam sekumpulan	
	dokumen	10
Gambar 2.4.	Arsitektur HDFS	11
Gambar 3.1.	Blok diagram penelitian	17
Gambar 4.1.	Blok Diagram Sistem	20
Gambar 4.2.	Peta Titik CCTV yang Diamati	21
Gambar 4.3.	Tampilan Masing - Masing CCTV yang Diamati.	21
Gambar 4.4.	Sketsa Titik CCTV yang Diamati	
Gambar 4.5.	Kendaraan k1 Terbaca pada CCTV 1	23
Gambar 4.6.	Kendaraan k1 Terbaca pada CCTV 2	
Gambar 4.7.	Kendaraan k1 Terbaca pada CCTV 3	24
Gambar 4.8.	Kendaraan k2 Terbaca pada CCTV 2	25
Gambar 4.9.	Kendaraan k2 Terbaca pada CCTV 3	25
Gambar 4.10.	Kendaraan k2 Terbaca pada CCTV 4	25
Gambar 4.11.	Kendaraan k2 Terbaca pada CCTV 5	
Gambar 4.12.	Kendaraan k3 Terbaca pada CCTV 2	26
Gambar 4.13	Kendaraan k4 Terbaca pada CCTV 2	
Gambar 4.14.	Kendaraan k4 Terbaca pada CCTV 5	27
Gambar 4.15.	Arsitertur Sistem	27
Gambar 4.16.	Grafik Model Paralel	28
Gambar 4.17.	Grafik Model Paralel	29
Gambar 4.18.	Grafik Model Paralel	29
Gambar 4.19.	Struktur Data Input untuk HDFS	
Gambar 4.20.	Flowchart fungsi mapper	
Gambar 4.21.	Flowchart fungsi reducer	37
Gambar 4.22.	Ilustrasi tahap 4.1.2.5. – 4.1.2.6	38
Gambar 4.23.	Class diagram	39
Gambar 4.24.	Pembuatan Direktori pada HDFS	40
Gambar 4.25.	Direktori pada HDFS	40
Gambar 4.26.	Input Data ke Dalam HDFS	41
Gambar 4.27.	Hasil Input Data pada HDFS	

Gambar 4.28.	Running Program MapReduce	45
Gambar 5.1.	Menampilkan Output dari Program	49
Gambar 5.2.	Output Program	49
Gambar 5.3.	Analisa Hasil Pengamatan	52

DAFTAR TABEL

Tabel 4.1. Rata – Rata Waktu Perpindahan antar CCTV	. 23
Tabel 4.2. Analisa Hasil Pengamatan	. 51



BAB I PENDAHULUAN

Pada bab ini akan dibahas mengenai latar belakang yang mendasari penulisan Tugas Akhir. Di dalamnya mencakup identifikasi permasalahan pada topik Tugas Akhir kemudian dirumuskan menjadi permasalahan yang selanjutnya diberikan batasan-batasan dalam pembahasan pada Tugas Akhir ini.

1.1. Latar Belakang

Dengan perkembangan sosial yang begitu cepat, industri transportasi juga menghadapi tantangan yang belum pernah dihadapi sebelumnya. Salah satunya adalah pertumbuhan jumlah kendaraan pribadi yang begitu pesat.[1] Ditambah lagi dengan tidak tersedianya sistem transportasi umum yang cukup memadai kebutuhan sosial masyarakatnya, maka tantangan tersebut semakin menjadi besar. Salah satu faktor penyebab sistem transportasi umum yang kurang memadai kebutuhan sosial masyarakat adalah penentuan rute atau trayek transportasi umum yang kurang sesuai dengan kebutuhan mobilitas masyarakat.

Maka dari itu, dibutuhkan suatu teknologi yang mampu merumuskan suatu sistem yang dapat mengatasi permasalahan penentuan rute transportasi umum. Dengan memanfaatkan *Closed Circuit Television* (CCTV) yang telah tersebar di beberapa ruas persimpangan jalan di Kota Surabaya, penulis telah melakukan pemantauan jalur perjalanan pada setiap kendaraan bermotor yang melewati beberapa titik CCTV di Kota Surabaya. Dikarenakan teknologi yang digunakan oleh CCTV tersebut masih belum memadai untuk membaca jalur perjalanan suatu kendaraan maupun hanya untuk membaca plat nomor suatu kendaraan secara otomatis, maka pengenalan kendaraanbeserta jalur perjalanannya dilakukan secara manual.

Pada Tugas Akhir ini, penulis telah mengimplementasikan big data pada data jalur perjalanan kendaraan yang lalu-lalang di ruas jalan beberapa CCTV yang telah dipantau sebelumnya. Menururt Badan Pusat Statistik, pada tahun 2015, terdapat 2.126.168 kendaraan bermotor yang berada di Kota Surabaya,[2] sehingga dibutuhkan suatu imlementasi big data yang dapat mengolah jumlah volume data yang begitu besar. Implementasi big data ini diterapkan pada ekosistem Hadoop, dimana HDFS sebagai sistem file untuk penyimpanan big data dan MapReduce sebagai model komputasi paralel. Karena data yang diolah cukup besar, maka peulis menggunakan model komputasi paralel untuk mengolah data tersebut, agar pemrosesan data dapat dilakukan secara jauh lebih efisien dibandingkan model komputasi biasa (sekuensial).

Diharapkan dengan diterapkannya big data pada sistem transportasi ini, dapat digunakan sebagai dasar bagi pemerintah Kota Surabaya dalam pengambilan keputusan untuk pembukaan rute transportasi umum.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang sudah diuraikan, pokok permasalahan yang dikaji dalam tugas akhir ini adalah :

- 1. Bagaimana implementasi data transportasi ke dalam ekosistem Hadoop?
- 2. Bagaimana mengolah data jalur perjalanan kendaraan bermotor pada ekosistem Hadoop?
- 3. Bagaimana menentukan urutan jalur perjalanan terbanyak yang dilewati oleh kendaraan bermotor?

1.3. Batasan Masalah

Penulisan Tugas Akhir ini difokuskan pada pembahasan dengan beberapa batasan masalah sebagai berikut :

- 1. Sistem transportasi yang diteliti adalah sistem transportasi di lima titik CCTV di Surabaya Pusat.
- 2. Kendaraan roda dua, angkutan kota, dan bus kota tidak diperhitungkan di dalam penelitian ini.
- 3. Framework atau platform yang digunakan pada penelitian ini adalah Hadoop.
- 4. Bahasa pemrograman yang digunakan adalah Java.

1.4. Tujuan

Berdasarkan rumusan masalah di atas, tujuan dari penulisan Tugas Akhir ini sebagai berikut :

- 1. Mengimplementasi data transportasi ke dalam ekosistem Hadoop.
- 2. Mengolah data jalur perjalanan kendaraan bermotor pada ekosistem Hadoop.
- 3. Menentukan urutan jalur perjalanan terbanyak yang dilewati oleh kendaraan bermotor.

1.5. Manfaat

Berdasarkan rumusan masalah di atas, tujuan dari penulisan Tugas Akhir ini sebagai berikut :

- 1. Memberikan pengetahuan bagi pembaca mengenai implementasi big data pada ekosistem Hadoop.
- 2. Memberikan masukan terhadap pemerintah Kota Surabaya, khususnya Dinas Perhubungan mengenai efektifitas rute transportasi umum Kota Surabaya.
- 3. Memberikan informasi untuk digunakan dalam riset selanjutnya.

1.6. Sistematika Penulisan Tugas Akhir

Sistematika dari penulisan Tugas Akhir ini adalah sebagai berikut:

1. BAB I PENDAHULUAN

Bab ini menjelaskan tentang gambaran umum dari penulisan Tugas Akhir ini yang meliputi latar belakang masalah, perumusan masalah, batasan masalah, tujuan, manfaat penelitian, dan sistematika penulisan.

2. BAB II TINJAUAN PUSTAKA

Bab ini berisi tentang materi-materi yang mendukung Tugas Akhir ini, antara lain big data, hadoop, MapReduce, HDFS, dan Kondisi Sistem Lalu Lintas Cerdas Kota Surabaya.

3. BAB III METODE PENELITIAN

Pada bab ini dibahas tentang langkah – langkah dan metode yang digunakan untuk menyelesaikan Tugas Akhir.

4. BAB IV PERANCANGAN DAN IMPLEMENTASI SISTEM Pada bab ini akan menguraikan bagaimana perancangan dan tahapan tahapan dalam implementasi sistem, yaitu.

5. BAB V HASIL UJI COBA DAN PEMBAHASAN

Bab ini menjelaskan mengenai hasil dari sistem yang telah dibuat disertai dengan penjelasan pembahasan.

6. BAB VI PENUTUP

Bab ini berisi kesimpulan yang diperoleh dari pembahasan masalah sebelunya serta saran yang diberikan untuk pengembangan selanjutnya.

BAB II TINJAUAN PUSTAKA

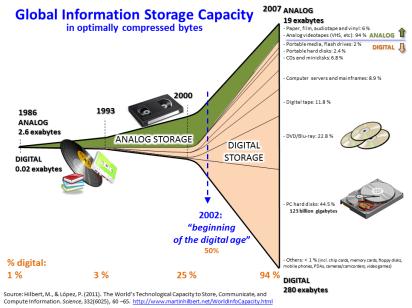
Pada bab ini dibahas mengenai dasar teori yang digunakan dalam penyusunan Tugas Akhir ini. Dasar teori yang dijelaskan adalah big data, Hadoop, MapReduce, HDFS, dan kondisi sistem lalu lintas cerdas Kota Surabaya.

2.1. Big Data

Big data adalah istilah yang digunakan untuk merujuk pada studi dan aplikasi dari himpunan data yang besar dan kompleks dimana aplikasi perangkat lunak pemrosesan data tradisional tidak memadai untuk berurusan dengan himpunan data tersebut. Tantangan big data meliputi penangkapan data, penyimpanan data, analisis data, pencarian, *sharing*, transfer, visualisasi, query, *updating*, privasi informasi dan sumber data. Terdapat beberapa konsep yang berkaitan dengan big data, semula terdapat tiga konsep, yaitu *volume*, *variety* (keragaman data), dan *velocity* (kecepatan pertumbuhan data).[3] konsep lain yang kemudian dikaitkan dengan big data adalah *veracity* (kebenaran atau ketelitian) dan *value* (nilai dari data itu sendiri).[4]

Belakangan ini, istilah big data cenderung merujuk pada penggunaan analisis prediktif, analisis tingkah laku user, atau metode-metode analisis data lanjutan tertentu yang mengekstrak nilai dari data, dan kadang juga diterapkan kepada ukuran himpunan data tertentu. Ada sedikit keraguan bahwa jumlah data yang tersedia sekarang memang besar, tetapi itu bukan karakteristik yang paling relevan dari ekosistem data baru ini. Analisis himpunan data dapat menemukan korelasi baru untuk melihat tren bisnis, mencegah penyakit, memerangi kejahatan dan sebagainya. Para ilmuwan, eksekutif bisnis, praktisi kedokteran, iklan, dan pemerintahan biasanya menemukan kesulitan dengan himpunan data yang besar di berbagai bidang termasuk pencarian Internet, *fintech*, informatika perkotaan, dan informatika bisnis. Para ilmuwan menemukan keterbatasan dalam pekerjaan *e-Science*, termasuk meteorologi,

konektromiks, simulasi fisika kompleks, biologi, dan penelitian lingkungan.[5]



Gambar 2.1. Pertumbuhan dan digitalisasi kapasitas penyimpanan informasi global

Sumber: http://www.martinhilbert.net/WorldInfoCapacity.html

Himpunan - himpunan data tumbuh dengan sangat cepat, sebagian karena himpunan - himpunan data tersebut semakin dikumpulkan oleh penginderaan informasi internet yang murah dan beragam dari berbagai gawai, seperti perangkat seluler, aerial, log perangkat lunak, kamera, mikrofon, pembaca Identfikasi Frekuensi Radio (*Radio-Frequency Identidication –RFId*), dan jaringan - jaringan sensor nirkabel.[6] Kapasitas per kapita teknologi dunia untuk menyimpan informasi telah meingkat dua kali lipat setiap 40 bulan sejak tahun 1980-an, pada 2012, setiap hari 2,5 exabytes (2,5 × 1018 byte) data dihasilkan.[7] Berdasarkan prediksi laporan IDC, volume data global akan tumbuh secara eksponensial dari 4,4 zettabyte menjadi 44 zettabytes antara 2013 dan 2020. Pada 2025,

IDC memprediksi akan ada 163 zettabytes data. Satu pertanyaan untuk perusahaan besar adalah menentukan siapa yang harus memiliki inisiatif big-data yang mempengaruhi seluruh organisasi.

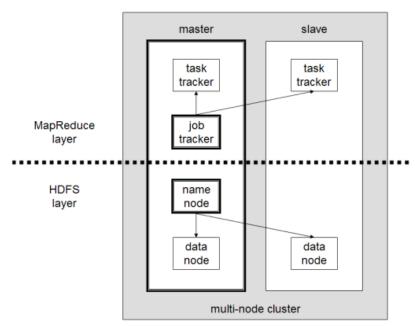
Sistem manajemen basis data relasional, statistik desktop, dan paket perangkat lunak untuk memvisualisasikan data sering mengalami kesulitan menangani big data. Pekerjaan tersebut mungkin memerlukan perangkat lunak paralel besar-besaran yang berjalan pada puluhan, ratusan, atau bahkan ribuan server. Apa yang dianggap sebagai "big data" bervariasi tergantung pada kemampuan *user* dan alat mereka, dan kemampuan yang semakin meluas membuat data besar menjadi target yang bergerak. Untuk beberapa organisasi, menghadapi ratusan gigabyte data untuk pertama kalinya dapat memicu kebutuhan untuk mempertimbangkan kembali opsi manajemen data. Bagi yang lain, mungkin diperlukan puluhan atau ratusan terabyte sebelum ukuran data menjadi pertimbangan yang signifikan.

2.2. Hadoop

Apache Hadoop adalah kumpulan utilitas perangkat lunak open source yang memfasilitasi penggunaan jaringan multi komputer untuk menyelesaikan masalah yang sejumlah besar data (atau biasa kita sebut dengan big data) dan komputasi.[8] Utilitas tersebut menyediakan kerangka kerja perangkat lunak untuk penyimpanan terdistribusi dan pengolahan big data menggunakan model pemrograman MapReduce. Awalnya dirancang untuk kelompok komputer yang dibangun dari perangkat keras yang biasa-biasa saja, meskipun tetap digunakan secara umum, namun Apache Hadoop juga telah digunakan pada kelompok perangkat keras kelas atas. Semua modul di Hadoop dirancang dengan asumsi mendasar bahwa kegagalan perangkat keras adalah kejadian umum dan harus ditangani secara otomatis oleh kerangka kerja.

Inti dari Apache Hadoop terdiri dari bagian penyimpanan, yang dikenal sebagai *Hadoop Distributed File System* (HDFS) atau jika diterjemahkan secara harfiah berarti Sistem File Terdistribusi Hadoop, dan bagian pemrosesan yang merupakan model pemrograman MapReduce. Hadoop membagi file menjadi blok-blok

besar dan mendistribusikannya di seluruh node dalam sebuah kelompok, kemudian mentransfer kode yang dikemas ke dalam node untuk memproses data secara paralel. Pendekatan ini mengambil keuntungan dari lokasi data, dimana node-node tersebut memanipulasi data yang mereka akses. Hal ini memungkinkan dataset untuk diproses lebih cepat dan lebih efisien daripada yang ada dalam arsitektur superkomputer yang lebih konvensional yang bergantung pada sistem file paralel di mana komputasi dan data didistribusikan melalui jaringan berkecepatan tinggi.



Gambar 2.2. Sebuah cluster multi-node Hadoop

Kerangka dasar Apache Hadoop terdiri dari modul-modul berikut:

1. *Hadoop Common* - berisi pustaka dan utilitas yang dibutuhkan oleh modul Hadoop lainnya,

- 2. *Hadoop Distributed File System* (HDFS) sistem file terdistribusi yang menyimpan data pada komputer biasa, menyediakan *bandwidth* yang sangat tinggi di seluruh kelompok,
- 3. *Hadoop YARN* diperkenalkan pada tahun 2012, adalah platform yang bertanggung jawab untuk mengelola sumber daya komputasi dalam setiap kelompok dan menggunakannya untuk menjadwalkan aplikasi para *user*,

Hadoop MapReduce - implementasi dari model pemrograman MapReduce untuk pemrosesan data yang berskala besar.

Istilah Hadoop tidak hanya merujuk pada modul dasar dan sub-modul yang disebutkan di atas, tetapi juga untuk ekosistem atau kumpulan paket perangkat lunak tambahan yang dapat dipasang 'pada' atau 'bersama' Hadoop, seperti Apache Pig, Apache Hive, Apache HBase, Apache Phoenix, Apache Spark, Apache ZooKeeper, Cloudera Impala, Apache Flume, Apache Sqoop, Apache Oozie, dan Apache Storm. Komponen MapReduce dan HDFS dari Apache Hadoop terinspirasi oleh paper Google di MapReduce dan Sistem File Google.

Kerangka Hadoop sendiri sebagian besar ditulis dalam bahasa pemrograman Java, dengan beberapa kode asli dalam C dan utilitas command line ditulis sebagai shell script. Meskipun kode MapReduce Java itu umum, namun bahasa pemrograman apa pun dapat digunakan dengan "Hadoop Streaming" untuk menerapkan bagian "map" dan "reduce" dari program user. Proyek lain dalam ekosistem Hadoop mengekspos User Interface yang lebih beragam.

2.2.1. MapReduce

MapReduce adalah sebuah model pemrograman dan implementasi berhubungan untuk dan yang memproses menghasilkan himpunan data yang besar dengan algoritma yang terdistribusi dan paralel pada sebuah kelompok data. Program MapReduce terdiri dari method 'map',yang melakukan penyaringan dan penyortiran (seperti menyortir siswa dengan nama depan ke dalam antrian, satu antrian untuk setiap nama), method 'reduce', yang melakukan operasi ringkasan (seperti menghitung jumlah siswa di setiap antrian, menghasilkan frekuensi nama). Kerangka kerja MapReduce mengatur pemrosesan dengan menyusun server yang terdistribusi, menjalankan berbagai tugas secara paralel, mengelola semua komunikasi dan transfer data antara berbagai bagian sistem, dan menyediakan redundansi dan toleransi kesalahan.

```
function map(String name, String document):
    // name: document name
    // document: document contents
    for each word w in document:
        emit (w, 1)

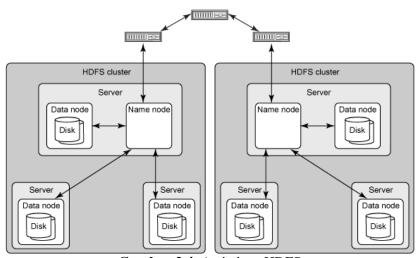
function reduce(String word, Iterator partialCounts):
    // word: a word
    // partialCounts: a list of aggregated partial counts
    sum = 0
    for each pc in partialCounts:
        sum += pc
    emit (word, sum)
```

Gambar 2.3. Contoh MapReduce kanonik menghitung kemunculan setiap kata dalam sekumpulan dokumen

Model ini adalah spesialisasi dari strategi split-apply-combine untuk analisis data.[9] Hal ini terinspirasi oleh funsi map dan reduce yang biasa digunakan dalam pemrograman fungsional, meskipun tuiuannya dalam kerangka MapReduce tidak sama dengan bentuk aslinya.[10] Kontribusi utama dari kerangka MapReduce bukanlah fungsi map dan reduce yang aktual, tetapi skalabilitas dan toleransi yang dicapai untuk berbagai aplikasi mengoptimalkan mesin eksekusi. Dengan demikian, implementasi MapReduce single threaded biasanya tidak akan lebih cepat daripada implementasi tradisional (non MapReduce), setiap keunggulan dari MapReduce biasanya hanya terlihat pada implementasi multi threaded. Penggunaan model ini bermanfaat hanya ketika operasi shuffle terdistribusi dioptimalkan) dan fitur toleransi kesalahan dari bekerja. Mengoptimalkan kerangka MapReduce ikut biava komunikasi sangat penting untuk algoritma MapReduce yang baik.[11]

Library dari MapReduce telah ditulis dalam banyak bahasa pemrograman, dengan tingkat pengoptimalan yang berbeda-beda. Implementasi open source yang populer yang mendukung shuffle terdistribusi adalah bagian dari Apache Hadoop. Istilah MapReduce awalnya hanya mengacu pada teknologi milik Google, tetapi sejak itu telah digeneralisasikan. Pada tahun 2014, Google tidak lagi menggunakan MapReduce sebagai model pemrosesan big data utama mereka, dan pengembangan pada Apache Mahout telah beralih ke mekanisme yang lebih kapabel dan mekanisme yang lebih tidak berorientasi pada disk yang menggabungkan kapabilitas map dan reduce secara utuh.

2.2.2. HDFS



Gambar 2.4. Arsitektur HDFS

HDFS adalah proyek Apache Software Foundation dan sub proyek dari proyek Apache Hadoop. Hadoop sangat ideal untuk menyimpan data dalam jumlah besar hingga ber-terabyte dan petabyte, dan menggunakan HDFS sebagai sistem penyimpanannya. HDFS memungkinkan Anda menghubungkan node (PC biasa) yang terdapat dalam kelompok di mana file data didistribusikan. Kemudian kita dapat mengakses dan menyimpan file data sebagai

satu sistem file yang mulus. Akses ke file data ditangani secara streaming, artinya aplikasi atau perintah dijalankan secara langsung menggunakan model pemrosesan MapReduce.

2.3. Implementasi Big Data pada ITS

Dewasa ini big data big data menjadi fokus penelitian dalam Sistem Transportasi Cerdas atau *Intelligent Transportation System (ITS)* yang dapat dilihat di banyak proyek di seluruh dunia.[12] *ITS* akan menghasilkan sejumlah besar data. Big data yang dihasilkan akan memiliki dampak besar pada desain dan penerapan *ITS* sehingga menjadi lebih aman, lebih efisien, dan menguntungkan.

Salah satu studi kasus aplikasi analisis big data dalam ITS adalah pembacaan jalur perjalanan kendaraan bermotor. Dari data perjalanan kendaraan bermotor tersebut, dapat dilakukan pengolahan data lebih lanjut, yaitu penentuan rute transportasi umum yang mengadopsi rute perjalanan kendaraan pribadi yang paling banyak terjadi.

2.4. Kondisi Sistem Lalu Lintas Cerdas Kota Surabaya

Surabaya, Sidoarjo, Mojokerto, Gresik, Bangkalan, Pasuruan, Lamongan dalam satu sistem tata ruang yang terintegrasi dalam struktur jaringan infrastruktur jalan, transportasi, komunikasi dan seterusnya. Integrasi sistem tata ruang dalam struktur jaringan infrastruktur jalan, transportasi, teknologi komunikasi dan informasi dan seterusnya menuju Greater Surabaya.

Juni 2012, Surabaya dinobatkan sebagai salah satu kota terbaik pada Peningkatan e-Government Indonesia (PeGi) 2012. Dinas Perhubungan kota Surabaya terus berupaya berkontribusi pada pemeringkatan PeGI dengan inisiatif memanfaatkan Teknologi Informasi dan Komunikasi dalam program-program pembangunan sistem cerdas (intelligent) untuk manajemen dan rekayasa Lalu-Lintas. Salah satu hasil program tersebut adalah penerapan awal Intelligent Transportation System (ITS). Pada tahap awal, sistem ini berupa sebuah Adaptive Traffic Control System sebagai upaya modernisasi ATCS konvensional yang telah dimiliki oleh Surabaya.

Upaya ini didukung oleh jaringan CCTV Surabaya sebagai bagian dari Traffic Management System. Intelligent Transport Sistem yaitu Sistem cerdas untuk mendukung manajemen transportasi dengan memanfaatkan teknologi (informasi, komunikasi, sensor, kontrol dan komputerisasi) untuk membangun sistem informasi dan manajemen transportasi secara otomatis. Adaptive Traffic Control System merupakan sistem yang mampu melakukan pengaturan waktu nyala lampu lalu-lintas (signal timing) secara real-time berdasarkn kondisi traffic saat itu, termasuk akibat keperluan (demand) khusus dan optimasi kapasitas arus lalu-lintas secara total.

Tujuan dari penerapan SITS ini antara lain adalah [13]:

- 1. Meningkatkan Keselamatan Lalu Lintas, diantaranya dengan cara mencegah/mengurangi kecelakan lalu lintas dan mengurangi kerusakan akibat kecelakaan.
- 2. Meningkatkan Kelancaran Lalu Lintas, diantaranya dengan cara mengoptimalkan siklus lampu lalu lintas baik secara otomatis maupun secara manual.
- 3. Menjaga kelestarian lingkungan dengan cara mengurangi polusi kendaraan akibat antrian kendaraan di ruas dan persimpangan. Dengan meningkatnya waktu tempuh dan berkurangnya waktu antrian dipersimpangan, diharapkan polusi kendaraan juga makin berkurang.

BAB III METODE PENELITIAN

Pada bab ini dijelaskan langkah-langkah yang digunakan dalam penyusunan Tugas Akhir. Disamping itu, dijelaskan pula prosedur dan proses pelaksanaan tiap-tiap langkah yang dilakukan dalam menyelesaikan Tugas Akhir.

3.1. Tahapan Penelitian

Pada penelitian tugas akhir ini dilakukan beberapa langkah sebagai berikut :

1. Studi Literatur

Pada tahap ini dilakukan identifikasi permasalahan dengan mencari referensi yang menunjang penelitian yang berupa Tugas Akhir, Jurnal Internasional, buku, maupun artikel yang berhubungan dengan topik Tugas Akhir ini.

2. Akuisisi Data

Akuisisi data ini dilakukan untuk mengumpulkan himpunan data kendaraan yang melalui CCTV yang terletak di kawasan Surabaya Pusat selama 30 menit. Lalu data tersebut dikonversi sehingga dapat diproses lebih lanjut.

3. Pembacaan Jalur Perjalanan Kendaraan pada MapReduce

Pada tahap ini dilakukan pembacaan jalur perjalanan kendaraan dengan cara membuat fungsi yang dapat membaca jalur perjalanan kendaraan dengan memanfaatkan data yang sudah didapat.

4. Pengolahan Data Jalur Perjalanan Kendaraan pada HDFS

Data jalur perjalanan kendaraan yang telah dibaca pada tahap sebelumnya diolah pada database yang berbasis HDFS.

5. Penyusunan Laporan

Pada tahap ini dilakukan penyusunan laporan berdasarkan pada hasil penelitian yang telah dilakukan.

3.2. Objek dan Aspek Penelitian

Objek yang digunakan pada penelitian ini adalah data kendaraan yang melalui CCTV di Kawasan Surabaya Pusat selama 30 menit kendaraan. Rekaman yang diamati adalah rekaman pada hari Senin, tanggal 25 November 2018 pukul 11.00 – 11.30. Sedangkan CCTV yang diamati adalah CCTV yang terletak pada:

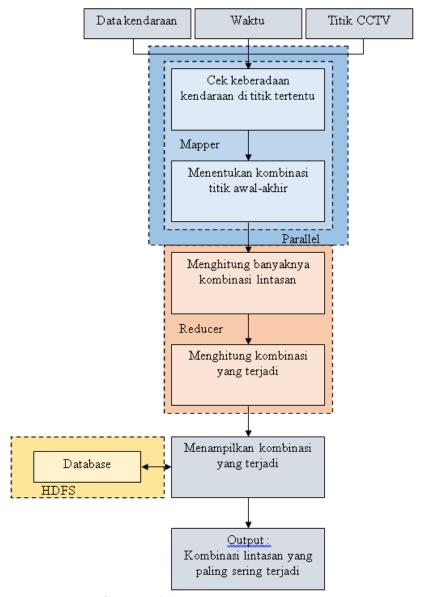
- 1. Jl. Kayoon Sonokembang
- 2. Jl. Panglima Sudirman
- 3. Jl. Urip Sumoharjo Pandegiling
- 4. Keputran Dinoyo
- 5. Jl. Raya Ngagel Sulawesi

Data penelitian diambil dengan cara mengamati setiap kendaraan yang terekam oleh 5 CCTV tersebut.

3.3. Blok Diagram

Data yang telah diperoleh dikonversi sehingga didapatkan himpunan data berupa data kendaraan dan lokasi dimana kendaraan tersebut terdeteksi (titik CCTV). Setelah itu, setiap kendaraan dicek keberadaannya di setiap titik, lalu ditentukan di titik mana kendaraan tersebut pertama dan terakhir kali terdeteksi. Langkah ini dilakukan oleh mapper dan dikerjakan secara parallel. Kemudian setelah pasangan titik awal-akhir dari diketahui setiap kendaraan. dihitunglah berapa kali pasangan tersebut terjadi. Langkah tersebut dilakukan oleh reducer. Setelah menghitung jumlah kejadian pasangan, data – data tersebut diolah dan disimpan di database yang berbasis HDFS, sekaligus ditampilkan hasilnya. Output yang ditampilkan berupa pasangan lintasan beserta banyaknya pasangan tersebut terjadi.

Untuk lebih ringkasnya, berikut adalah blok diagram yang digunakan untuk mengolah data yang telah diperoleh :



Gambar 3.1. Blok diagram penelitian

BAB IV PERANCANGAN DAN IMPLEMENTASI SISTEM

Bab ini menjelaskan tentang perancangan yang diperlukan dan implementasi sistem.

4.1. Perancangan Sistem

Pada bagian ini, akan dijelaskan tentang perancangan sistem yang meliputi perancangan data, perancagnan blok diagram sistem, serta perancangan *class* diagram.

4.1.1. Perancangan Data

Data-data yang digunakan dalam program ini dapat dibedakan menjadi dua jenis, yaitu data input dan data output.

1. Data Input

Data input adalah data yang digunakan sebagai masukan (input) dari program. Input ini yang kemudian diolah oleh program melalui tahap tahap tertentusehingga menghasilkan keluaran (output) yang diinginkan. Data input yang digunakan adalah data kendaraan bermotor yang melewati setiap CCTV selama 30 menit.

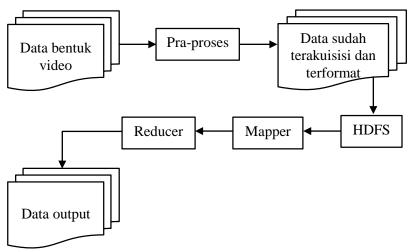
2. Data Output

Data output merupakan data yang dihasilkan oleh program setelah proses-proses tertentu selesai dilakukan. Data keluaran pada program ini berupa data pasangan titik awal-akhir perjalanan kendaraan bermotor beserta banyaknya pasangan tersebut dilalui. Kemudian dari data output tersebut akan dianalisa bagaimana preferensi jalur perjalanan kendaraan bermotor di lima titik CCTV tersebut.

4.1.2. Perancangan Blok Diagram Sistem

Pada Gambar 4.1 di bawah dijelaskan tentang tahapan-tahapan yang dilakukan untuk mencapai tujuan Tugas Akhir. Dalam tugas akhir ini *tools* untuk komputasi menggunakan bahasa pemrograman java pada ekosistem Hadoop dengan menggunakan NetBeans IDE

8.2. Program yang dibuat dalam NetBeans IDE 8.2 disimpan dengan nama *VehicleRoute*. Terdapat dua *class* yang keduanya akan disimpan dalam *package* default, yaitu *Source Packages*.



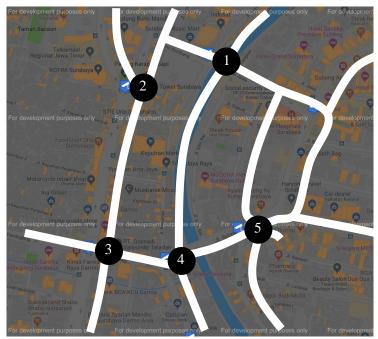
Gambar 4.1. Blok Diagram Sistem

4.1.2.1. Akuisisi Data (Pra-Proses)

Pada tahap pra-proses, dilakukan pengkonversian data secara manual, yaitu pengenalan kendaraan yang terekam oleh kamera CCTV, dicatat dengan memperhatikan waktu, letak CCTV, dan data kendaraan. Letak CCTV sendiri akan dinotasikan sebagai berikut:

- 1. Jl. Kayoon Sonokembang sebagai CCTV 1
- 2. Jl. Panglima Sudirman sebagai CCTV 2
- 3. Jl. Urip Sumoharjo Pandegiling sebagai CCTV 3
- 4. Keputran Dinoyo sebagai CCTV 4
- 5. Jl. Raya Ngagel Sulawesi sebagai CCTV 5

Gambar 4.2 adalah peta jalur perjalanan kendaraan bermotor dan titik – titik CCTV yang diamati. Sedangkan 8 adalah tampilan setiap titik CCTV yang diamati.

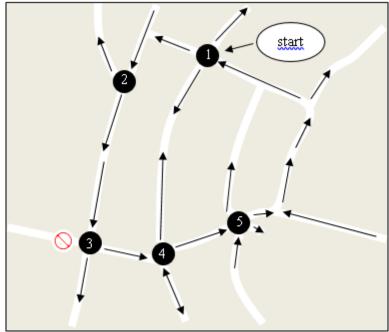


Gambar 4.2. Peta Titik CCTV yang Diamati



Gambar 4.3. Tampilan Masing - Masing CCTV yang Diamati

Gambar 4.3 merupakan tampilan visual dari setiap titik CCTV yang diamati dalam penelitian ini. Sedangkan berikut ini adalah sketsa dari titik – titik CCTV tersebut beserta arah arus lalu lintas di berbagai ruas simpangannya.

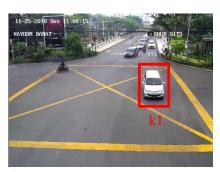


Gambar 4.4. Sketsa Titik CCTV yang Diamati

Sketsa arah arus lalu lintas pada Gambar 4.4 telah disesuaikan dengan sudut pandang kamera CCTV yang terpasang di setiap titik.

Proses akusisi data akan menghasilkan data yang dapat digunakan sebagai data input yang dapat disimpan pada HDFS dan nantinya dapat diolah oleh program mapreduce.

Berikut akan dilakukan simulasi proses pengakuisisian data sehingga didapat data input yang dapat disimpan pada HDFS dan kemudian dapat diolah oleh program MapReduce.



Gambar 4.5. Kendaraan k1 Terbaca pada CCTV 1

Pertama, setiap kendaraan yang terekam oleh setiap CCTV diamati dengan cara diberi ID dan dicatat waktu terekamnya dan titik CCTV yang merekam kendaraan vang diamati tersebut. Sebagai contoh. kendaraan pada gambar disamping diberi ID = k1.

dengan waktu = 00.15 (waktu yang dicatat hanya bagian menit dan detiknya saja), dan CCTV yang merekam adalah CCTV 1.

Berikut ini adalah tabel yang menunjukkan jarak antar CCTV dan berdasarkan data pada Lampiran 1, didapatkan pula rata – rata waktu yang dibutuhkan untuk berpidah antar CCTV. Dengan demikian, kita dapat menghitung rata – rata kecepatan kendaraan yang melewati setiap segmen antar CCTV.

Tabel 4.1. Rata - Rata Waktu Perpindahan Kendaraan antar CCTV

Titik CCTV	Jarak	Waktu	Kecepatan
1 - 2	350 m	183 s	6,84 km/h
2 - 3	500 m	43 s	41,76 km/h
3 – 4	280 m	39 s	25,92 km/h
4 - 5	300 m	131 s	8,28 km/h

Berdasarkan tabel di atas, dapat kita simpulkan bahwa rata – rata waktu yang dibutuhkan setiap kendaraan untuk berpindah antar CCTV kurang dari 5 menit, sehingga jika terdapat kendaraan yang waktu perpindahan antar CCTV lebih dari 5 menit, maka kendaraan tersebut akan dianggap sebagai kendaraan baru dan akan diberi ID yang berbeda dengan ID sebelumnya.

Dalam proses akuisisi data, akan dijumpai banyak kemungkinan kasus kendaraan yang terekam oleh sistem CCTV, di antaranya :

 Kasus pertama adalah kasus dimana kendaraan melewati beberapa CCTV secara berurutan dengan rentang waktu yang dibutuhkan untuk berpindah antar CCTV kurang dari 5 menit. Berikut adalah contoh kendaraan yang mengalami kasus pertama.



Gambar 4.6. Kendaraan k1 terbaca pada CCTV 2



Gambar 4.7 Kendaraan k1 Terbaca pada CCTV 3

Pada Gambar 4.6 dan 4.7, kendaraan dengan ID k1 melewati CCTV 2 menuju CCTV 3 dengan rentang waktu kurang dari 5 menit. Sehingga k1 dikatakan sebagai kendaraan yang mengalami kasus pertama.

2. Kasus kedua adalah kasus dimana kendaraan melewati beberapa CCTV secara berurutan selayaknya pada kasus pertama, namun pada kasus kedua ini, kendaraan berpindah antar CCTV membutuhkan waktu lebih dari 5 menit. Berikut adalah contoh kendaraan yang mengalami kasus kedua.



Gambar 4.8. Kendaraan k2 Terbaca pada CCTV 2



Gambar 4.9. Kendaraan k2 Terbaca pada CCTV 3



Gambar 4.10. Kendaraan k2 Terbaca pada CCTV 4



Gambar 4.11. Kendaraan k2 Terbaca pada CCTV 5

Pada Gambar 4.8 sampai 4.11, kendaraan dengan ID k2 melewati CCTV 1 menuju CCTV 2 dengan rentang waktu kurang dari 5 menit, namun dari CCTV 2 menuju CCTV 3, rentang waktu yang dibutuhkan lebih dari 5 menit. Sehingga k2 dikatakan sebagai kendaraan yang mengalami kasus kedua.

3. Kasus ketiga adalah kasus dimana kendaraan hanya melewati satu titik CCTV. Berikut adalah contoh kendaraan yang mengalami kasus ketiga.



Gambar 4.12. Kendaraan k3 Terbaca pada CCTV 2

Pada gambar di atas, kendaraan dengan ID k3 hanya melewati CCTV 2 saja dan tidak pernah terbaca lagi di CCTV berikutnya. Sehingga k3 dikatakan sebagai kendaraan yang mengalami kasus ketiga.

4. Kasus keempat adalah kasus terakhir, yaitu kasus diaman suatu kendaraan melewati beberapa CCTV tidak secara berurutan. Berikut adalah contoh kendaraan yang mengalami kasus ketiga.



Gambar 4.13. Kendaraan k4 Terbaca pada CCTV 2

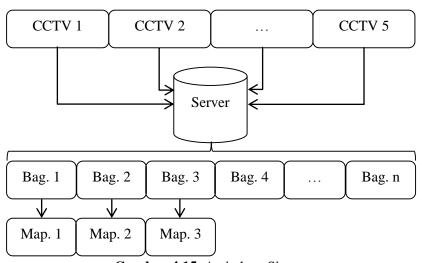


Gambar 4.14. Kendaraan k4 Terbaca pada CCTV 5

Pada gambar di atas, kendaraan dengan ID k4 melewati CCTV 2, mengikuti alur lalu-lintas pada sistem, seharusnya kendaraan tersebut akan terbaca di CCTV 3 lalu kemudian di CCTV 4, namun, kendaraan tersebut malah terbaca pada CCTV 5. Sehingga k4 dikatakan sebagai kendaraan yang mengalami kasus keempat.

4.1.2.2. Arsitektur Sistem Secara Umum

Berikut adalah gambaran tentang sistem yang akan dibuat :



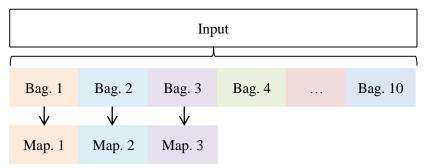
Gambar 4.15. Arsitektur Sistem

Arsitektur sistem secara umum idealnya mengambil nilai input dari setiap CCTV yang diamati. Setiap CCTV mengirim data kendaraan yang terbaca pada CCTV tersebut setiap satuan waktu tertentu kepada server. Selanjutnya server mengatur data yang diterima menjadi bagian – bagian yang akan diproses oleh fungsi *mapper* secara paralel. Karena prasarana yang belum memadai untuk menerapkan sistem yang ideal, maka tugas dari server yang mengatur data input menjadi bagian – bagian yang akan diproses oleh fungsi mapper dilakukan secara manual.

4.1.2.3. Perancangan Model Paralel

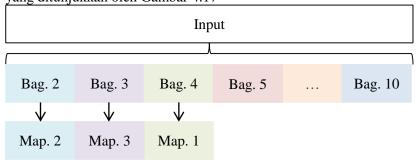
Pada tahap ini, akan dirancang model yang akan mengolah data input secara paralel. Diasumsikan setiap CCTV mengirim data input kepada server selama 10 satuan waktu, maka server akan mengatur data tersebut menjadi 10 bagian, dimana setiap bagian akan diolah oleh fungsi *mapper*. Terdapat tiga fungsi *mapper* yang akan mengolah setiap bagian data input tersebut yang dilakukan secara paralel. Sehingga jika salah satu fungsi *mapper* telah selesai mengolah suatu bagian data input, maka fungsi tersebut akan langsung mengambil bagian data input selanjutnya untuk diproses.

Gambar 4.16 menunjukkan bagaimana proses paralel dilakukan oleh fungsi mapper, dimana bagian input pertama hingga bagian input ketiga diproses terlebih dahulu oleh ketiga fungsi mapper.



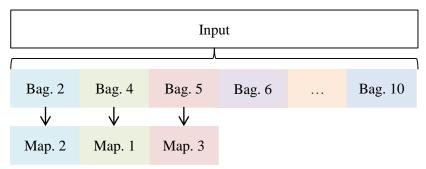
Gambar 4.16. Grafik Model Paralel

Dimisalkan *mapper* pertama telah terlebih dahulu selesai mengolah bagian input pertama, maka selanjutnya fungsi mapper pertama akan mengambil bagian input keempat untuk diolah seperti yang ditunjukkan oleh Gambar 4.17



Gambar 4.17. Grafik Model Paralel

Lalu dimisalkan lagi mapper ketiga selesai mengolah bagian input ketiga, maka fungsi mapper ketiga akan mengambil bagian selanjutnya, yakni bagian kelima untuk diolah seperti yang ditunjukkan pada Gambar 4.18, dan proses seperti ini dilakukan hingga semua bagian input diolah.



Gambar 4.18. Grafik Model Paralel

Jika terdapat input baru yang akan diolah, maka input baru tersebut akan ditempatkan setelah bagian input 10, dengan kata lain setiap input baru akan ditempatkan pada akhir antrean.

Untuk menetapkan banyaknya fungsi mapper yang akan bekerja, maka pada mapred-site yang terdapat pada folder etc/hadoop

harus ditambahkan sebuah property yang menyatakan bahwa banyaknya fungsi reduce yang bekerja adalah sebanyak tiga.

Berikut adalah property untuk menentukan banyaknya fungsi mapper yang akan bekerja.

4.1.2.4. Perancangan Struktur Data dan Penyimpanan Data pada HDFS

HDFS adalah suatu sistem file terdistribusi, yaitu sistem file yang menyimpan data tidak hanya dalam satu media penyimpanan. Disinilah nantinya data yang telah diakuisisi sebelumnya akan diinputkan, lalu data tersebut akan diolah oleh program mapreduce.

Karena data yang akan diinputkan ke dalam HDFS haruslah dapat diolah oleh fungsi *mapper*, sehingga data hasil akuisisi akan disimpan dengan struktur data seperti pada Gambar 4.19. Data tersebut disimpan dalam *array of String* satu dimensi, agar proses *parsing* oleh fungsi *mapper* dapat dilakukan lebih mudah.

0102 2 k1 0123 3 k1 0012 2 k2 0102 3 k2 0647 4 k2 0740 5 k2 0824 2 k3 0534 2 k4 1019 5 k4

Gambar 4.19. Struktur Data Input untuk HDFS

Pada Gambar 4.19, index ke- i pada array tersebut menunjukkan waktu suatu kendaraan terbaca pada CCTV, index ke-i+1 menunjukkan titik CCTV yang membaca kendaraan tersebut, dan index ke- i+2 adalah ID kendaraan yang terbaca.

Output dari data yang diolah secara otomatis akan tersimpan pada HDFS dengan direktori yang telah dibuat pada saat menjalankan program.

4.1.2.5. Perancangan Fungsi Mapper

Pada tahap ini, fungsi *mapper* akan memetakan pasangan nilai input menjadi sautu himpunan pasangan — pasangan nilai intermediate. Pada kasus ini, nilai input yang dimaksud berupa pasangan waktu suatu kendaraan terbaca oleh CCTV, lokasi CCTV yang membaca kendaraa tersebut, dan ID kendaraan yang terbaca. Sedangkan output dari fungsi *mapper* ini, yaitu pasangan — pasangan nilai intermediate, adalah titik dimana suatu kendaraan pertama kali terbaca oleh CCTV dan titik terakhir kendaraan tersebut terbaca.

Fungsi *mapper* akan mengolah bagian inputnya sehingga menghasilkan output yang dapat diolah oleh fungsi reducer, yaitu pasangan CCTV – CCTV dimana suatu kendaraan pertama kali terbaca dan terakhir kali terbaca oleh CCTV. Misalkan kendaran A terbaca pada CCTV X pada T_1 lalu terbaca lagi pada CCTV Y pada waktu T_2 , maka jalur perjalanan kendaraan A tersebut akan disimpan dalam bentuk String sebagai "X – Y".

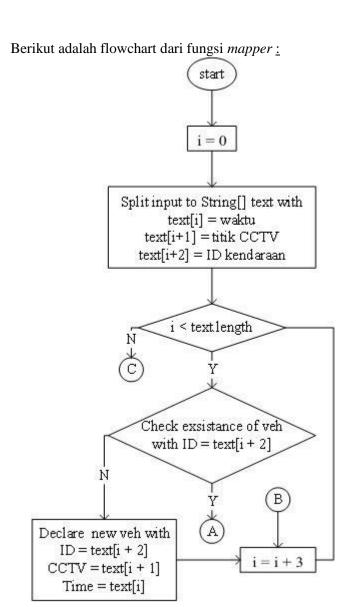
Berdasarkan contoh kejadian yang terlihat di pada Gambar 4.19, fungsi *mapper* akan mambaca jalur perjalanan dari setiap kendaraan yang terekam, sehingga didapat kendaraan k1 bepergian dari CCTV 2 ke CCTV 3, kendaraan k2 terbaca sebagai dua kendaraan berbeda yaitu kendaraan yang bepergian dari CCTV 2 ke CCTV 3, dan kendaraan yang bepergian dari CCTV 4 ke CCTV 5, hal ini dikarenakan rentang waktu pada perjalanan dari CCTV 2 menuju CCTV 3 lebih dari 5 menit, sehingga pada CCTV 3, k2 terbaca sebagai kendaraan baru lagi. Hasil *mapping* untuk kendaraan k3 tidak ditampilkan karena kendaraan k3 hanya terbaca pada satu CCTV saja. Lalu kendaraan k4 terbaca dari CCTV 2 ke CCTV 5.

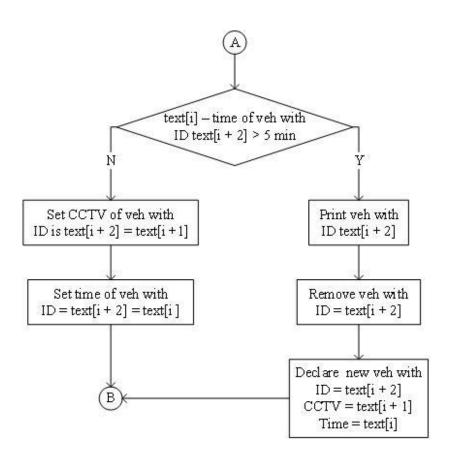
Untuk dapat diolah, data input terlebih dahulu dijadikan suatu array, damana array tersebut memuat waktu kendaraan terbaca oleh CCTV pada index ke- i (yang selanjutnya disebut "waktu"), CCTV yang membaca kendaraan tersebut pada index ke- i+2 (yang selanjutnya disebut "CCTV"), dan ID kendaraan yang terbaca pada index ke- i+3 (yang selanjutnya disebut "ID"), dimana i adalah bilangan asli dari 0 sampai 99.

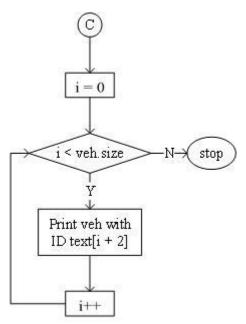
Setelah data input disimpan pada array, selanjutnya array tersebut diproses dengan cara dicek apakah kendaraan dengan ID =

elemen array pada index ke- i+2 terdapat pada himpunan kendaraan yang sudah dicek (telah terbaca oleh suatu CCTV) atau belum. Jika sudah terdapat pada himpunan tersebut, dicek kembali apakah kendaraan tersebut terbaca pada CCTV sebelumnya lebih dari 5 menit yang lalu atau tidak. Jika kendaraan tersebut terbaca pada CCTV sebelumnya lebih dari 5 menit, maka kendaraan tersebut dinyatakan sebagai kendaraan yang baru terbaca, sehingga kendaraan tersebut dideklarasikan pada himpunan tersebut dengan atribut waktu = elemen pada array dengan index ke- i, CCTV = elemen pada array dengan index ke- i+1, dan ID = elemen pada array dengan index kei+2. Jika ternyata kendaraan tersebut terbaca oleh CCTV sebelumnya kurang dari 5 menit yang lalu, maka kendaraan tersbut hanya mengupdate atribut waktu-nya menjadi elemen array pada index ke- I dan menambahkan atribut CCTV dengan elemen array pada index kei+1, sehingga nantinya setiap kendaraan yang dicek akan memiliki himpunan CCTV yang pernah dilaluinya. Sama seperti kasus pertama, jika kendaraan dengan ID = elem array ke- i+2 belum sama sekali pernah terbaca pada CCTV manapun, maka kendaraan tersebut akan dideklarasikan pada himpunan kendaraan yang sudah terbaca sebagai suatu anggota baru dengan atribut waktu = elemen pada array dengan index ke- i, CCTV = elemen pada array dengan index ke-i+1, dan ID = elemen pada array dengan index ke-i+2.

Selanjutnya setiap kendaraan yang telah dicek tersebut akan diambil dua lokasi atau titik CCTV pada himpunan CCTV yang pernah dilaluinya, yaitu elemen pertama pada himpunan CCTV tersebut dan dinyatakan sebagai titik awal kendaraan tersebut terbaca atau memasuki sistem transportasi yang diteliti, dan elemen terakhir himpunan pada himpunan CCTV tersebut yang akan dinyatakan sebagai titik akhir kendaraan tersebut terbaca atau titik dimana kendaraan tersebut keluar dari sistem transportasi yang diteliti. Yang pada akhirnya, fungsi *mapper* ini menghasilkan output berupa pasangan – pasangan titik awal dan titik akhir dari setiap kendaraan yang terbaca oleh sistem.







Gambar 4.20. Flowchart fungsi mapper

Flowchart di atas menerangkan bagaimana suatu bagian input yang berisi 100 kendaraan diolah pada fungsi *mapper*.

Langkah – langkah di atas juga dapat ditulis dengan *pseudocode* sebagai berikut :

```
String[] text = split input;
   ArrayList<veh> vehicle;
2.
3. For i = 0 as i < text.length with <math>i = i+3 {
4.
      If checkExistance of vehicle with ID = text[i+2] {
5.
         If text[i] - time of vehicle with ID = text[i+2] <</pre>
          5 min {
6.
             If first and last CCTV of vehicle with ID =
             text[i+2] different {
7.
                Print (the first CCTV and the last CCTV);
8.
9.
            Remove vehicle with ID = text[i+2];
10.
            Add new vehicle with ID = text[i+2], CCTV =
            text[i+1], time = [i];
11.
12.
        Else {
13.
            add CCTV of the vehicle = text[i+1];
14.
             add time of the vehicle = text[i];
15.
16.
     Else add new vehicle with ID = text[i+2], CCTV =
      text[i+1], time = [i];
17. }
18. For i = 0 as i < vehicle.size with i++ {
     If first and last CCTV of vehicle with ID = text[i+2]
      different {
20.
         Print (the first CCTV and the last CCTV);
21.
22. }
```

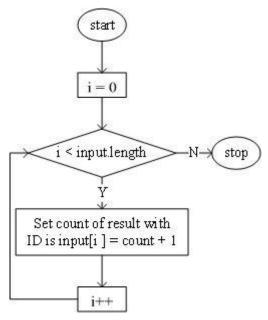
4.1.2.6. Perancangan Fungsi Reducer

Setelah pasangan nilai input diproses menjadi pasangan nilai intermediate oleh fungsi *mapper*, selanjutnya pasangan nilai intermediate tersebut diolah oleh fungsi *reducer* menjadi nilai output final. Pada kasus ini, pasangan titik awal dan titik akhir suatu kendaraan akan diolah oleh fungsi *reducer*, yaitu dengan cara dihitung frekuensi kemunculan dari setiap pasangan titik awal dan titik akhir. Output dari fungsi *reducer* ini adalah pasangan titik awal dan titik akhir perjalanan setiap kendaraan beserta frekuensi kemunculan dari setiap pasangan tersebut.

Misal terdapat n perjalanan kendaraan yang telah terbaca pada fungsi mapper, maka dari n perjalanan tersebut akan dihitung berapa banyak perjalanan yang melalui jalur " $X_k - Y_k$ ". Sehingga nantinya akan terhitung jumlah kejadian setiap pasangan jalur perjalanan yang telah terbaca pada fungsi mapper.

Berdasarkan contoh kejadian pada Gambar 4.19, fungsi reducer akan menghitung setiap pasangan jalur perjalanan kendaraan yang telah dibaca oleh fungsi mapper, yaitu jalur perjalanan kendaraan dari CCTV 2 menuju CCTV 3 sebanyak dua kali, dan masing – masing satu kali untuk jalur perjalanan dari CCTV 4 menuju CCTV 5 dan dari CCTV 2 menuju CCTV 5.

Berikut adalah flowchart dari fungsi reducer:



Gambar 4.21. Flowchart fungsi reducer

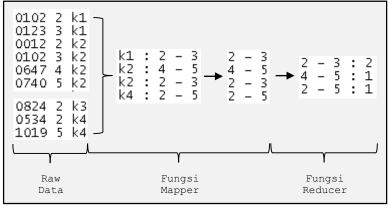
Flowchart di atas menjelaskan bagaimana fungsi *reducer* mengolah data pasangan data intermediate yang dihasilkan oleh fungsi *mapper*. Output dari fungsi mapper yang berupa himpunan pasangan titik awal dan titik akhir dari suatu kendaraan yang

melewati sistem adalah suatu input bagi fungsi *reducer* ini. Setiap pasangan titik awal dan titik akhir dihitung jumlah kejadiannya, sehingga fungsi ini menghasilkan output berupa himpunan pasangan titik awal dan titik akhir beserta dengan jumlah kejadian pasangan tersebut.

Langkah – langkah di atas juga dapat ditulis dengan *pseudocode* sebagai berikut :

```
1. IntWritable result;
2. For i = 0 as i < input.length with i++ {
3. Set count of result with ID = input[i] = count+1;
4. }</pre>
```

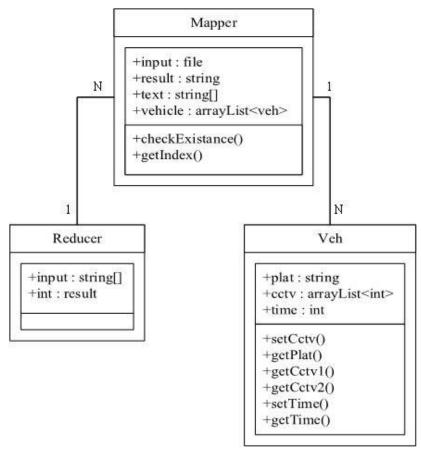
Gambar 4.22 menunjukkan ilustrasi bagaimana fungsi *mapper* mengolah *raw* data sehingga menjadi suatu output yang selanjutnya dapat diolah oleh fungsi *reducer* menjadi *final* output. Pada ilustrasi terserbut, maksud dari hasil pembacaan dari fungsi mapper "k1:2-3" kendaraan k1 bepergian dari CCTV 2 menuju CCTV 3, dan maksud dari "2-3" adalah terdapat sebuah kendaraan yang bepergian dari CCTV 2 menuju CCTV 3. Lalu hasil penghitungan oleh fungsi reducer yang bertuliskan "2-3:2" berarti bahwa jumlah kendaraan yang melalui CCTV 2 menuju CCTV 3 sebanyak dua buah kendaraan.



Gambar 4.22. Ilustrasi tahap 4.1.2.5. - 4.1.2.6.

4.1.3. Perancangan Class Diagram

Pada class diagram di bawah ini akan dijelaskan tentang UML yang akan dijadikan dasar sebagai pembuatan program MapReduce. Program yang akan dibuat nantinya memiliki tiga kelas, yakni *mapper*, *reducer*, dan veh, yang memiliki atribut dan method seperti yang terlihat pada diagram di berikut.



Gambar 4.23. Class diagram

4.2. Implementasi Sistem

Pada bagian ini dijelaskan tentang implementasi struktur data dan penyimpanan data pada HDFS, implementasi model paralel pada fungsi *mapper*, dan implementasi fungsi *reducer*.

4.2.1. Implementasi Struktur Data dan Penyimpanan Data pada HDFS

Hasil dari akuisisi data dicatat dengan struktur data seperti pada Subbab 4.1.2.1.. Sebelum menginput data hasil akuisisi tersebut ke dalam HDFS, terlebih dahulu dibuat direktori pada HDFS dimana data tersebut akan diinputkan. Berikut adalah sintaks untuk membuat direktori pada HDFS:

C:\Users\John>hdfs dfs -mkdir -p /user/input Gambar 4.24. Pembuatan Direktori pada HDFS

Gambar di atas menunjukkan bahwa direktori "input" akan dibuat pada super direktori "user", dimana super direktori ini dibuat pada PC "John". Berikut adalah tampilan direktori sebelum data diinputkan :



Gambar 4.25. Direktori pada HDFS

Setelah direktori tersedia, data input telah dapat diinputkan ke dalam HDFS dengan cara sebagai berikut :

```
G:\Users\John>hdfs dfs -put C:\Users\John\Desktop\input1.txt /user/input
C:\Users\John>hdfs dfs -put C:\Users\John\Desktop\input2.txt /user/input
C:\Users\John>hdfs dfs -put C:\Users\John\Desktop\input3.txt /user/input
C:\Users\John>hdfs dfs -put C:\Users\John\Desktop\input4.txt /user/input
C:\Users\John>hdfs dfs -put C:\Users\John\Desktop\input5.txt /user/input
C:\Users\John>hdfs dfs -put C:\Users\John\Desktop\input6.txt /user/input
C:\Users\John>hdfs dfs -put C:\Users\John\Desktop\input7.txt /user/input
C:\Users\John>hdfs dfs -put C:\Users\John\Desktop\input8.txt /user/input
C:\Users\John>hdfs dfs -put C:\Users\John\Desktop\input8.txt /user/input
C:\Users\John>hdfs dfs -put C:\Users\John\Desktop\input9.txt /user/input
C:\Users\John>hdfs dfs -put C:\Users\John\Desktop\input9.txt /user/input
C:\Users\John>hdfs dfs -put C:\Users\John\Desktop\input10.txt /user/input
```

Gambar 4.26. Input Data ke Dalam HDFS

Gambar di atas menunjukkan bahwa file input1.txt — input10.txt yang berada pada desktop pada PC "John" telah berhasil diinputkan ke dalam HDFS dengan direktori "input" yang berada pada super direktori "user". Berikut adalah tampilan direktori /user/input setelah data berhasil diinput :

iser/input								G
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
-144-1	John	supergroup	2.33 KB	1/7/2019, 10:37:26 PM	1	128 MB	input1.bt	
-rw-rr	John	supergroup	2.1 KB	1/7/2019, 10:38:59 PM	1	128 MB	input10.bt	
-144-1	John	supergroup	2.19 KB	1/7/2019, 10:37:35 PM	1	128 MB	input2.bt	
- FW-FF	John	supergroup	2.44 KB	1/7/2019, 10:37:42 PM	1	128 MB	input3.bt	
-rw-rr	John	supergroup	2.32 KB	1/7/2019, 10:38:09 PM	1	128 MB	input4.bt	
-144-11	John	supergroup	2.68 KB	1/7/2019, 10:38:18 PM	1	128 MB	input5.bt	
-TW-TT	John	supergroup	2.84 KB	1/7/2019, 10:38:25 PM	1	128 MB	input6.bt	
- F F	John	supergroup	2.73 KB	1/7/2019, 10:38:32 PM	1	128 MB	input7.bt	
-TW-TT	John	supergroup	2.18 KB	1/7/2019, 10:38:43 PM	1	128 MB	input8.txt	
-rw-rr	John	supergroup	2.53 KB	1/7/2019, 10:38:50 PM	1	128 MB	input9.bt	

Gambar 4.27. Hasil Input Data pada HDFS

4.2.2. Implementasi Fungsi Mapper

Data yang sebelumnya telah diinputkan ke HDFS kemudian diolah oleh fungsi *mapper* untuk dijadikan kumpulan String, String

tersebut adalah pasangan titik awal dan titik akhir jalur perjalanan suatu kendaraan bermotor.

Pada tahap ini, dibuat suatu subclass dari class Mapper, yang diberi nama "MyMapper". Class MyMapper hanya memiliki satu atribut, yaitu "word", yang merepresentasikan pasangan titik awalakhir yang akan dihitung frekuensi kejadiannya. Class MyMapper memiliki method "map" yang digunakan untuk mengolah data input menjadi pasangan-pasangan titik awal dan titik akhir jalur perjalanan suatu kendaraan dalam bentuk String.

Berikut adalah implementasi dari fungsi *mapper* :

```
public static class MyMapper extends Mapper <LongWritable,
Text, Text, IntWritable> {
   private Text word = new Text();
   public void map(LongWritable key, Text value, Context
   context) throws IOException, InterruptedException {
      String[] stringArr = value.toString().split("\\s+");
     ArrayList<veh> sortedVeh = new ArrayList<veh>();
      for (int i = 0; i < stringArr.length; i = i + 3) {
         if (checkExistence(sortedVeh, stringArr[i + 2])) {
            if ((Integer.parseInt(stringArr[i]) -
            sortedVeh.get(getIndex(sortedVeh, stringArr[i +
            2])).getTime()) > 500) {
               if (sortedVeh.get(i).getCctv1() !=
               sortedVeh.get(i).getCctv2()) {
                  word.set(sortedVeh.get(getIndex(sortedVeh
                  , stringArr[i + 2])).getCctv1() + " - " +
                  sortedVeh.get(getIndex(sortedVeh,
                  stringArr[i + 2])).getCctv2());
                 context.write(word, new IntWritable(1));
            sortedVeh.remove(getIndex(sortedVeh,
            stringArr[i + 2]));
            sortedVeh.add(new veh(stringArr[i + 2],
            Integer.parseInt(stringArr[i + 1]),
            Integer.parseInt(stringArr[i])));
         } else {
            sortedVeh.get(getIndex(sortedVeh, stringArr[i +
            2])).setCctv(Integer.parseInt(stringArr[i +
            1]));
            sortedVeh.get(getIndex(sortedVeh, stringArr[i +
            2])).setTime(Integer.parseInt(stringArr[i]));
         } else {
            sortedVeh.add(new veh(stringArr[i + 2],
            Integer.parseInt(stringArr[i + 1]),
            Integer.parseInt(stringArr[i])));
      for (int i = 0; i < sortedVeh.size(); i++) {
         if (sortedVeh.get(i).getCctv1() !=
         sortedVeh.get(i).getCctv2()) {
            word.set(sortedVeh.get(i).getCctv1() + " - "
            sortedVeh.get(i).getCctv2());
            context.write(word, new IntWritable(1));
     }
   }
```

Pada class "MyMapper", terdapat dua method yang dipanggil, yaitu :

1. checkExistence(ArrayList<veh> x, String y)

Method ini digunakan untuk mengecek apakah suatu String y (merepresentasikan plat nomor) dimiliki oleh salah satu objek di ArrayList x (merepresentasikan kumpulan data kendaraan) atau tidak. Berikut adalah implementasi dari method checkExistence(ArrayList<veh> x, String y).

```
public static boolean checkExistence(ArrayList<veh> x,
String y) {
   boolean flag = false;
   for (int i = 0; i < x.size(); i++) {
      if (x.get(i).getPlat().equals(y)) {
        flag = true;
        break;
      }
   }
   return flag;
}</pre>
```

2. getIndex(ArrayList<veh> x, String y)

Method ini digunakan untuk mencari index dari suatu objek di ArrayList x (kumpulan kendaraan) yang memiliki atribut String y (plat nomor). Berikut adalah implementasi dari method getIndex(ArrayList<veh> x, String y)

```
public static int getIndex(ArrayList<veh> x, String y) {
   int flag = 0;
   for (int i = 0; i < x.size(); i++) {
      if (x.get(i).getPlat().equals(y)) {
        flag = i;
        break;
      }
   }
   return flag;
}</pre>
```

4.2.3. Implementasi Fungsi Reducer

Setelah mendapatkan pasangan titik awal dan titik akhir dalam bentuk String, dilakukan penghitungan frekuensi kemunculan pasangan – pasangan tersebut oleh fungsi *reducer*. Berikut adalah implementasi dari fungsi *reducer*:

```
public static class MyReducer extends Reducer<Text,
IntWritable, Text, IntWritable> {
   private IntWritable result = new IntWritable();
   public void reduce(Text key, Iterable<IntWritable>
   values, Context context)
   throws IOException, InterruptedException {
     int sum = 0;
     for (IntWritable val : values) {
        sum += val.get();
     }
     result.set(sum);
     context.write(key, result);
}
```

"MyReducer" merupakan subclass dari class Reducer, yang memiliki atribut "result", yaitu atribut yang merepresentasikan banyaknya kejadian untuk setiap pasangan yang ada. Selain itu, class ini juga memiliki method "reduce" yang berfungsi untuk menghitung banyaknya kejadian dari suatu pasangan titik awal-akhir.

4.2.4. Running Program

Setelah data terinput ke dalam HDFS dan fungsi *mapper* dan *reducer* telah terimplementasi, maka program MapReduce siap untuk di-running. Berikut adalah cara running program MapReduce pada ekosistem Hadoop:

C:\Users\John>hadoop jar C:\Users\John\Desktop\VehicleRoute.jar VehicleRoute /us er/input /user/output

Gambar 4.28. Running Program MapReduce

Gambar di atas menunjukka bahwa program MapReduce yang pada kasus ini disimpan dengan nama "VehicleRoute" yang

tersimpan pada Desktop pada PC "John" yang memiliki *main class* mengeksekusi data input yang berada pada direktori /user/input dan nantinya hasil dari eksekusi oleh program ini disimpan pada direktori /user/output. Berikut merupakan proses *running* dari program MapReduce:

```
19/01/10 13:41:38 INFO client.RMProxy: Connecting to ResourceManager at localhos t/12?.0.0.1:8032
19/01/10 13:41:39 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
19/01/10 13:41:40 INFO mapreduce.JobSubmitter: number of splits:10
19/01/10 13:41:40 INFO mapreduce.JobSubmitter: number of splits:10
19/01/10 13:41:40 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_15
47/102344447.0001
19/01/10 13:41:41 INFO impl.YarnClientImpl: Submitted application application_15
47/10234447.0001
19/01/10 13:41:41 INFO mapreduce.Job: The url to track the job: http://10.212.1.
56:8088/proxy/application_1547102344447_0001/
19/01/10 13:41:51 INFO mapreduce.Job: Running job: job_1547102344447_0001
19/01/10 13:41:51 INFO mapreduce.Job: bob job_1547102344447_0001 running in uber mode: false
19/01/10 13:41:51 INFO mapreduce.Job: map 0% reduce 0%
19/01/10 13:42:07 INFO mapreduce.Job: map 0% reduce 0%
19/01/10 13:42:15 INFO mapreduce.Job: map 90% reduce 0%
19/01/10 13:42:20 INFO mapreduce.Job: map 100% reduce 0%
19/01/10 13:42:21 INFO mapreduce.Job: map 100% reduce 0%
19/01/10 13:42:22 INFO mapreduce.Job: Job job_1547102344447_0001 completed succe ssfully
19/01/10 13:42:21 INFO mapreduce.Job: Job Job_1371-0.

19/01/10 13:42:22 INFO mapreduce.Job: Counters: 49

File System Counters

File: Number of bytes read=10038

FILE: Number of bytes written=1377024

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes written=47

HDFS: Number of bytes written=47

HDFS: Number of read operations=3

HDFS: Number of large read operations=3

HDFS: Number of write operations=3

HDFS: Number of write operations=2

Job Counters
                                                                                                HDFS: Number of write operations=2

Job Counters

Launched map tasks=10

Launched reduce tasks=1

Data-local map tasks=10

Total time spent by all maps in occupied slots (ms)=64194

Total time spent by all reduces in occupied slots (ms)=5759

Total time spent by all reduces in occupied slots (ms)=5759

Total time spent by all reduces tasks (ms)=5759

Total time spent by all reduce tasks (ms)=5759

Total voore-milliseconds taken by all map tasks=64194

Total voore-milliseconds taken by all reduce tasks=5759

Total megabyte-milliseconds taken by all map tasks=65734656

Total megabyte-milliseconds taken by all reduce tasks=5897216

Map-Reduce Framework

Map input records=10
                                                                                                                                                                                         Total megasyce miliseconds taken by all reduces Framework
Map input records=18
Map output bytes=8360
Map output bytes=8360
Map output bytes=1881
Combine input records=0
Combine input records=0
Reduce input groups=5
Reduce input groups=5
Reduce input records=836
Reduce input records=836
Reduce output records=5
Spilled Records=1672
Spifled Maps =10
Failed Shuffles=0
Merged Map outputs=10
GC time elapsed (ms)=1673
CPU time spent (ms)=5280
Dirital memory (bytes) snapshot=2916249600
Virtual memory (bytes) snapshot=2240806912
Errors
Dan In=0
                                                                                              IOTAL COMMISSION Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
URONG_REDUCE=0
WRONG_REDUCE=0
WRONG_REDUCE=0
                                                                                                    File Input Format Counters
Bytes Read=24096
File Output Format Counters
Bytes Written=47
```

BAB V HASIL UJI COBA DAN PEMBAHASAN

Pada bab ini dijelaskan tentang hasil dan pembahasan dari program yang telah dibuat.

5.1. Hasil Akuisisi Data

Pada penelitian ini, data yang digunakan adalah data rekaman CCTV di lima lokasi yang telah disebutkan sebelumnya. Data tersebut didapatkan dari Dinas Perhubungan Kota Surabaya pada tanggal 26 November 2018. Hasil akusisi data dari video menghasilkan input sebanyak 1000 kendaraan, yang dapat dilihat pada Lampiran 1.

5.2. Hasil Uji Coba Sistem

Program yang telah dijalankan sebelumnya menghasilkan data pasangan titik awal dan titik akhir dari jalur perjalanan kendaraan-kendaraan yang melewati setiap CCTV yang diamatai beserta frekuensi kejadian pasangan – pasangan tersebut. Hasil running program akan menghasilkan ouput yang disimpan pada direktori yang telah disebutkan pada saat running, yaitu /user/output dengan nama "part-r-00000". Berikut adalah cara menampilkan data output yang terdapat pada HDFS:

C:\Users\John>hdfs dfs -cat /user/output/part-r-00000 Gambar 5. Menampilkan Output dari Program

Sedangkan output dari running program tersebut adalah sebagai berikut :

1 -	2 -	3	32
2 -	3	635	
2 -	3 -	4	16
2 -	3 -	4 - 5	20
4 -	5	132	

Gambar 6. Output Program

Gambar 5.2 adalah hasil dari pengolahan data input pada HDFS oleh program MapReduce.

5.3. Analisa Hasil Pengolahan Data

Berdasarkan hasil yang telah didapatkan sebelumnya, maka dapat dianalisa bahwa sebanyak 635 kendaraan memiliki jalur perjalanan yang berawal di CCTV 2 dan berakhir di CCTV 3, yang berarti 635 kendaraan tersebut memasuki area pengamatan melalui Jl. Panglima Sudirman dan keluar dari pengamatan melalui Jl. Urip Sumoharjo. Pasangan jalur perjalanan kendaraan ini adalah pasangan titik awal dan titik akhir yang paling banyak dilalui oleh kendaraan. Sementara pasangan titik awal dan titik akhir yang paling sedikit dilalui adalah pasangan yang berawal dari Jl. Panglima Sudirman menuju Keputran. Selain itu, dapat dilihat bahwa kendaraan hanya memasuki sistem transportasi melalui tiga titik, yaitu Jl. Kayoon, Jl. Panglima Sudirman, dan Keputran, yang berarti setiap kendaraan yang melewati Jl. Urip Sumoharjo pasti juga melewati Jl. Panglima Sudirman, dan setiap kendaraan yang melewati Jl. Raya Ngagel juga pasti melewati Keputran. Hal ini karena tidak ada kendaraan yang pertama kali terbaca pada Jl. Urip Sumoharjo ataupun Jl. Raya Ngagel. Hal ini juga sesuai dengan sketsa sistem transportasi yang telah dibuat pada BAB IV.

Total jumlah kendaraan pada data input adalah 1000 kendaraan, namun pada akhirnya hanya ada 835, sehingga terdapat sebanyak 165 kendaraan yang tadinya terdapat pada data input, namun tidak ditampilkan pada output. Hal ini diakibatkan karena beberapa kendaraan mengalami proses akuisisi seperti kasus 2 dan kasus 3. Dimana kasus 2 adalah kendaraan yang melewati beberapa CCTV dengan rentang waktu lebih dari 5 menit, sehingga satu kendaraan yang sama akan terbaca oleh program sebanyak dua kali. Namun masalah utamanya bukan pada kendaraan yang mengalami proses akuisisi seperti kasus 2, melainkan kendaraan – kendaraan yang mengalami proses akuisisi seperti kasus 3, yaitu setiap kendaraan hanya melewati satu CCTV, sehingga kendaraan tersebut tidak diproses. Selisih antara kendaraan yang mengalami proses akusisi seperti kasus 3 dengan kendaraan yang mengalami proses

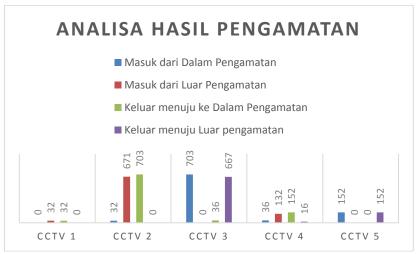
akuisisi seperti kasus 2 inilah data input yang tidak ditampilkan pada output, yaitu sebanyak 164 kendaraan.

Dari total 835 kendaraan tersebut, dapat dianalisa bahwa sebanyak 32 kendaraan masuk ke pengamatan melalui Jl. Kayoon, 671 kendaraan masuk ke pengamatan melalui Jl. Panglima sudirman, dan 132 lainnya masuk melalui keputran. Dari 32 kendaraan yang masuk ke Jl. Kayoon, seluruhnya akan menuju Jl. Panglima sudirman, sehingga jumlah kendaraan yang keluar dari Jl. Panglima sudirman beriumlah 703 kendaraan, dimana dari 703 kendaraan tersebut seluruhnya akan menuju Jl. Urip sumoharjo. Jl. Urip sumoharjo tidak menerima kendaraan dari luar pengamatan, yang berarti, sebanyak 703 kendaraan yang melewati Jl. Urip sumoharjo dipastikan adalah kendaraan yang melewati Jl. Panglima sudirman terlebih dahulu. Dari 703 kendaraan yang keluar dari Jl. Urip sumoharjo, 36 kendaraan diantaranya menuju keputran dan 667 kendaraan lainnya keluar dari pengamatan. Dikarenakan kendaraan yang keluar dari Jl. Urip sumoharjo menuju keputran, maka jumlah kendaraan yang menuju keputran sebanyak 169 kendaraan. Dari 169 kendaraan tersebut, sebanyak 152 kendaraan menuju Jl. Raya ngagel dan sisanya sebanyak 16 kendaraan keluar dari pengamatan. Seluruh kendaraan yang menuju cctv adalah kendaraan yang sebelumnya telah melewati keputran. Akhirnya. seluruh kendaraan yang keluar dari Jl. Raya ngagel akan keluar dari pengamatan. Analisa di atas dapat dirangkum dalam bentuk tabel seperti yang tertera berikut:

Tabel 5.1. Analisa Hasil Pengamatan

	Masu	k dari	Keluar ke		
CCTV	Dalam	Luar	Dalam	Luar	
	Pengamatan	Pengamatan	Pengamatan	Pengamatan	
1	-	32	32	-	
2	32	671	703	-	
3	703	-	36	667	
4	36	132	152	16	
5	152	-	-	152	

Selain dapat disajikan dalam bentuk tabel, hasil analisa tersebut juga dapat disajikan dalam bentuk diagram sebagai berikut :



Gambar 7. Analisa Hasil Pengamatan

BAB VI PENUTUP

Pada bab ini berisi tentang beberapa kesimpulan yang dihasilkan berdasarkan penelitian yang telah dilaksanakan dan saran yang dapat digunakan jika penelitian ini dikembangkan.

6.1. Kesimpulan

Berdasarkan analisis terhadap hasil pengujian program, maka dapat diambil kesimpulan sebagai berikut :

- 1. Data input yang berhasil diakuisisi dari data video yang diperoleh dari Dinas Perhubungan Kota Surabaya pada hari Senin tanggal 26 November 2018 selama 30 menit pada pukul 11.00 11.30 sebanyak 1000 kendaraan. Data tersebut lalu dikonversi menjadi *array of String* 1D sehingga dapat diproses oleh program MapReduce.
- 2. Data input yang telah terkonversi diimplementasikan ke dalam ekosistem Hadoop dengan cara diinputkan ke HDFS dan diproses oleh program MapReduce.
- 3. Jalur kendaraan yang paling banyak dialui oleh kendaraan adalah jalur perjalanan yang berawal dari Jl. Panglima sudirman dan berakhir pada Jl. Urip sumoharjo. Sehingga jika pemerintah kota surabaya akan membuka rute atau trayek transportasi umum dalam kota, maka sebaiknya pemerintah kota surabaya membuka rute yang berawal dari Jl. Panglima sudirman dan berakhir di Jl. Urip sumoharjo.

6.2. Saran

Ada beberapa hal yang penulis sarankan untuk pengembangan penelitian selanjutnya :

- 1. Kualitas gambar CCTV ditingkatkan agar setiap kendaraan yang melewati CCTV tersebut dapat terlihat plat nomornya.
- 2. Terapkan teknologi pembacaan plat nomor secara otomatis sehingga proses akuisisi data menjadi jauh lebih cepat.

3. Jika kedua hal di atas telah diterapkan, maka dapat dilakukan penelitian tentang sistem transportasi Kota Surabaya secara keseluruhan, tidak menggunakan *sampling*, sehingga hasil dari penelitian tersebut dapat langsung diimplementasikan sesuai kebutuhan sistem transportasi Kota Surabaya.

DAFTAR PUSTAKA

- [1] Zeng, Gang (2015). "Application of Big Data in Intelligent Traffic System". IOSR Journal of Computer Engineering (IOSR-JCE). 17 (1).
- [2] https://surabayakota.bps.go.id/statictable (dikunjungi pada 7 Desember 2018).
- [3] Laney, Doug (2001). "3D data management: Controlling data volume, velocity and variety". META Group Research Note. 6 (70).
- [4] Goes, Paulo B. (2014). "Design science research in top information systems journals". MIS Quarterly: Management Information Systems. 38 (1)
- [5] Reichman, O.J.; Jones, M.B.; Schildhauer, M.P. (2011). "Challenges and Opportunities of Open Data in Ecology". Science. 331 (6018): 703–5.
- [6] Segaran, Toby; Hammerbacher, Jeff (2009). Beautiful Data: The Stories Behind Elegant Data Solutions. O'Reilly Media. p. 257.
- [7] Hilbert, Martin; López, Priscila (2011). "The World's Technological Capacity to Store, Communicate, and Compute Information". Science. 332 (6025): 60–65.
- [8] http://hadoop.apache.org/releases.html (dikunjungi pada 20 September 2018)
- [9] Wickham, Hadley (2011). "The split-apply-combine strategy for data analysis". Journal of Statistical Software. 40: 1–29.
- [10] Lämmel, R. (2008). "Google's Map Reduce programming model Revisited". Science of Computer Programming. 70: 1–30.
- [11] Ullman, J. D. (2012). "Designing good MapReduce algorithms". XRDS: Crossroads, The ACM Magazine for Students. Association for Computing Machinery. 19: 30.
- [12] Zhu, Li; Yu, Fei R.; Wang, Yige; Ning, Bin; Tang, Tao (2018). "Big Data Analytics in Intelligent Transportation Systems: A Survey".
- [13] http://sits.dishub.surabaya.go.id/ver2/tentang-sits (dikunjungi pada 21 Agustus 2018).

LAMPIRAN

1. Hasil Akuisisi Data

```
k545 1449 2 k546 1450 2 k547 1451 2 k548 1452 2 k549 1453 2
k550 1454 2 k551 1455 2 k552 1456 2 k553 1457 2 k554 1459 2
k555 1501 2 k556 1503 2 k557 1505 2 k558 1506 2 k559 1507 2
k560 1559 3 k545 1600 3 k546 1601 3 k547 1602 3 k548 1603 3
k549 1604 3 k550 1605 3 k551 1606 3 k552 1607 3 k553 1608 3
k554 1609 3 k555 1610 3 k556 1611 3 k557 1612 3 k558 1612 3
k559 1614 3 k560 1722 4 k560 1803 5 k560 0951 4 k561 0952 4
k562 0953 4 k563 0954 4 k564 0955 4 k565 0957 4 k566 0959 4
k567 1001 4 k569 1003 4 k570 1005 4 k571 1015 4 k572 1030 4
k573 1049 4 k574 1149 5 k561 1153 5 k568 1157 5 k574 1143 4
k575 1144 4 k576 1145 4 k577 1146 4 k578 1147 4 k579 1149 4
k580 1151 4 k581 1153 4 k582 1155 4 k583 1157 4 k584 1200 4
k585 1203 4 k586 1206 4 k587 1209 4 k588 1212 4 k589 1216 4
k590 1220 4 k591 1224 4 k592 1228 4 k593 1232 4 k594 1252 4
k595 1307 4 k596 1318 4 k597 1411 5 k575 1412 5 k576 1413 5
k577 1414 5 k578 1415 5 k579 1417 5 k580 1419 5 k581 1421 5
k582 1423 5 k583 1425 5 k584 1428 5 k585 1431 5 k586 1434 5
k587 1437 5 k588 1440 5 k589 1444 5 k590 1448 5 k591 1449 5
k592 1450 5 k593 1451 5 k594 1452 5 k595 1453 5 k596 1454 5
k597 1343 4 k598 1345 4 k599 1346 4 k600 1347 4 k601 1348 4
k602 1349 4 k603 1351 4 k604 1353 4 k605 1355 4 k606 1357 4
k607 1359 4 k608 1402 4 k609 1405 4 k610 1408 4 k611 1411 4
k612 1414 4 k613 1418 4 k614 1422 4 k615 1426 4 k616 1430 4
k617 1434 4 k618 1436 4 k619 1438 4 k620 1439 4 k621 1549 5
k610 1555 5 k611 1601 5 k612 1607 5 k613 1613 5 k614 1624 5
k615 1635 5 k616 1646 5 k617 1657 5 k618 1708 5 k619 1728 5
k620 1743 5 k621 1509 1 k622 1807 2 k622 1840 3 k622 1611 1
k623 1912 2 k623 2050 3 k623 1709 1 k624 1712 1 k625 1716 1
k626 2016 2 k625 2022 2 k626 2106 3 k625 2113 3 k626 1819 1
k627 2125 2 k627 2213 3 k627 1929 1 k628 1934 1 k629 1937 1
k630 1941 1 k631 2243 2 k629 2246 2 k628 2246 2 k630 2422 3
k630 2426 3 k628 2426 3 k629 1508 2 k630 1509 2 k631 1510 2
k632 1511 2 k633 1512 2 k634 1514 2 k635 1516 2 k636 1518 2
k637 1520 2 k638 1522 2 k639 1525 2 k640 1528 2 k641 1531 2
k642 1534 2 k643 1537 2 k644 1539 2 k645 1541 2 k646 1543 2
k647 1545 2 k648 1547 2 k649 1548 2 k650 1549 2 k651 1550 2
k652 1551 2 k653 1552 2 k654 1553 2 k655 1554 2 k656 1555 2
```

```
k825 1938 5 k826 1943 5 k827 1948 5 k828 1953 5 k829 1957 5
k830 1958 4 k831 2000 4 k832 2013 4 k833 2015 4 k834 2019 4
k835 2025 4 k836 2031 4 k837 2033 4 k838 2039 4 k839 2043 4
k840 2053 5 k831 2054 5 k832 2055 5 k833 2056 5 k834 2057 5
k835 2102 5 k836 2107 5 k837 2112 5 k838 2117 5 k839 2118 5
k840 2103 1 k841 2151 1 k842 2201 1 k843 2239 1 k844 2544 2
k844 2551 3 k844 2422 1 k845 2441 1 k846 2716 2 k845 2744 2
k846 2837 3 k845 2904 3 k846 2017 2 k847 2018 2 k848 2019 2
k849 2020 2 k850 2022 2 k851 2024 2 k852 2026 2 k853 2028 2
k854 2030 2 k855 2033 2 k856 2036 2 k857 2039 2 k858 2104 2
k859 2221 3 k847 2222 3 k848 2223 3 k849 2224 3 k850 2225 3
k851 2227 3 k852 2229 3 k853 2231 3 k854 2233 3 k855 2235 3
k856 2238 3 k857 2241 3 k858 2245 3 k859 2357 4 k859 2536 5
k859 2106 2 k860 2107 2 k861 2108 2 k862 2109 2 k863 2110 2
k864 2112 2 k865 2114 2 k866 2116 2 k867 2118 2 k868 2120 2
k869 2123 2 k870 2126 2 k871 2129 2 k872 2132 2 k873 2135 2
k874 2139 2 k875 2143 2 k876 2147 2 k877 2151 2 k878 2155 2
k879 2156 2 k880 2157 2 k881 2158 2 k882 2159 2 k883 2200 2
k884 2201 2 k885 2201 2 k886 2202 2 k887 2202 2 k888 2203 2
k889 2203 2 k890 2204 2 k891 2204 2 k892 2205 2 k893 2205 2
k894 2246 3 k860 2247 3 k861 2248 3 k862 2249 3 k863 2250 3
k864 2252 3 k865 2254 3 k866 2256 3 k867 2258 3 k868 2300 3
k869 2303 3 k870 2306 3 k871 2309 3 k872 2312 3 k873 2315 3
k874 2319 3 k875 2323 3 k876 2327 3 k877 2331 3 k878 2335 3
k879 2340 3 k880 2345 3 k881 2350 3 k882 2355 3 k883 2400 3
k884 2401 3 k885 2402 3 k886 2403 3 k887 2404 3 k888 2405 3
k889 2406 3 k890 2406 3 k891 2407 3 k892 2407 3 k893 2407 3
k894 2633 4 k894 2207 2 k895 2208 2 k896 2209 2 k897 2210 2
k898 2212 2 k899 2214 2 k900 2216 2 k901 2218 2 k902 2220 2
k903 2223 2 k904 2226 2 k905 2229 2 k906 2232 2 k907 2235 2
k908 2239 2 k909 2243 2 k910 2247 2 k911 2251 2 k912 2255 2
k913 2300 2 k914 2305 2 k915 2310 2 k916 2311 2 k917 2312 2
k918 2313 2 k919 2314 2 k920 2315 2 k921 2316 2 k922 2316 2
k923 2317 2 k924 2317 2 k925 2318 2 k926 2318 2 k927 2318 2
k928 2318 2 k929 2318 2 k930 2408 3 k895 2409 3 k896 2410 3
k897 2412 3 k898 2414 3 k899 2416 3 k900 2418 3 k901 2420 3
k902 2423 3 k903 2426 3 k904 2429 3 k905 2432 3 k906 2435 3
```

BIODATA PENULIS



Penulis dengan sapaan Fachry dan nama lengkap Mohammad Fian Fachry Alfahmi ini bertempat tinggal di Kabupaten Situbondo. Anak pertama dari tiga bersaudara ini lahir pada tanggal 12 Mei 1996. Selama menjalani masa perkuliahan di Departemen Matematika ITS. penulis aktif sebagai anggota UKM Technopreneurship Development Center (TDC ITS)pada tahun pertama. Pada tahun kedua, penulis melanjutkan kiprahnya di UKM TDC ITS sebagai Entrepreneurship Development. Di asisten manajer

tahun ketiganya, penulis sempat "vakum" untuk berkuliah di lingkungan kampus ITS untuk mengikuti program *Study Exchange* selama dua semester (sepuluh bulan) di Catalunya, Spanyol. Kampus tempat penulis mengikuti program *study exchange* adalah *Universitat Rovira I Virgili* dan mengambil jurusan *Finance and Accounting*. Pada tahun keempat, penulis kembali melanjutkan perkuliahan di ITS untuk fokus mengejar ketertinggalan yang disebabkan oleh *study exchange* tersebut.

Tugas Akhir ini tak lepas dari kritik dan saran. Jika ada yang ingin didiskusikan ataupun ditanyakan, jangan ragu untuk menghubungi penulis via email : fian.fachry.fifa@gmail.com . Terimakasih dan semoga Tugas Akhir ini bermanfaat.