

**PEMANFAATAN APACHE KAFKA DAN SPARK
STRUCTURED STREAMING DALAM MONITORING JARINGAN:
STUDI KASUS DETEKSISERANGAN DDOS**

SKRIPSI

Oleh:

**FEBRIANTO KUDADIRI
NIM. 221113626**



**PROGRAM STUDI S-1 TEKNIK INFORMATIKA
FAKULTAS INFORMATIKA
UNIVERSITAS MIKROSKIL
MEDAN
2026**

UTILIZING APACHE KAFKA AND SPARK STRUCTURED STREAMING IN NETWORK MONITORING: A CASE STUDY OF DDOS ATTACK DETECTION

UNDERGRADUATE THESIS

By:

**FEBRIANTO KUDADIRI
ID NUMBER. 221113626**



**UNDERGRADUATE PROGRAM OF S-1 COMPUTER SCIENCE
FACULTY OF INFORMATICS
UNIVERSITAS MIKROSKIL
MEDAN
2026**

LEMBARAN PENGESAHAN

PEMANFAATAN APACHE KAFKA DAN SPARK STRUCTURED STREAMING DALAM MONITORING JARINGAN: STUDI KASUS DETEKSI SERANGAN DDOS

SKRIPSI

Diajukan untuk Melengkapi Persyaratan Guna
Mendapatkan Gelar Sarjana
Program Studi S-1 Teknik Informatika

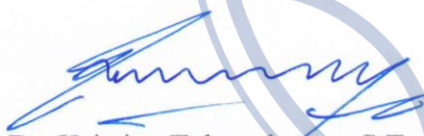
Oleh:

FEBRIANTO KUDADIRI
NIM. 221113626

Disetujui Oleh:

Dosen Pembimbing I,

Dosen Pembimbing II,


Dr. Kristian Telaumbanua, S.T., M.T.


Apriyanto Halim, S.Kom., M.Kom.

Medan, 17 Juli 2026
Diketahui dan Disahkan Oleh:

Ketua Program Studi
S-1 Teknik Informatika,


Carles Juliandy, S.Kom., M.Kom.

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Teknik Informatika Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : FEBRIANTO KUDADIRI

NIM : 221113626

Saya telah melaksanakan penelitian dan penulisan Skripsi dengan judul dan tempat penelitian sebagai berikut:

Judul Skripsi : Pemanfaatan Apache Kafka dan Spark Structured Streaming dalam Monitoring Jaringan: Studi Kasus Deteksi Serangan DDoS

Tempat Penelitian : -

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Skripsi saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Skripsi saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 28 April 2026

Saya yang membuat pernyataan,



FEBRIANTO KUDADIRI

Pemanfaatan Apache Kafka dan Spark Structured Streaming dalam Monitoring Jaringan: Studi Kasus Deteksi Serangan DDoS

Abstrak

Serangan Distributed Denial of Service (DDoS) menghasilkan lalu lintas jaringan dalam jumlah besar sehingga memerlukan sistem deteksi yang mampu memproses data secara cepat dan berkelanjutan. Penelitian ini bertujuan membangun pipeline distributed stream processing berbasis Apache Kafka, Spark Structured Streaming, dan Random Forest untuk mendukung deteksi DDoS secara near real-time. Apache Kafka digunakan sebagai message broker, Spark Structured Streaming sebagai engine pemrosesan data streaming, dan Random Forest sebagai model klasifikasi serangan. Evaluasi dilakukan menggunakan variasi input rate sebesar 100, 200, dan 400 record per detik dengan mengukur performa Kafka Producer, Spark Structured Streaming, model Random Forest, serta pipeline end-to-end berdasarkan metrik latency, throughput, dan kualitas klasifikasi. Hasil penelitian menunjukkan bahwa pipeline mampu memberikan respon deteksi secara near real-time hingga input rate 200 record per detik dengan mayoritas latency pemrosesan berada di bawah SLA 5 detik setelah fase inisialisasi. Model Random Forest mempertahankan akurasi di atas 99% pada seluruh skenario pengujian. Penurunan performa pada input rate 400 record per detik disebabkan oleh overhead Spark Structured Streaming dan akumulasi antrean data. Hasil penelitian menunjukkan bahwa arsitektur yang dikembangkan layak diterapkan untuk deteksi DDoS pada beban rendah hingga menengah serta menjadi dasar pengembangan sistem stream processing yang lebih skalabel.

Kata kunci: Apache Kafka, Spark Structured Streaming, Random Forest, DDoS, Distributed Stream Processing

Abstract

Distributed Denial of Service (DDoS) attacks generate massive network traffic, requiring detection systems capable of processing data rapidly and continuously. This study develops a distributed stream processing pipeline using Apache Kafka, Spark Structured Streaming, and the Random Forest algorithm to support near real-time DDoS detection. Apache Kafka functions as the message broker, Spark Structured Streaming as the streaming processing engine, and Random Forest as the attack classification model. The pipeline is evaluated under input rates of 100, 200, and 400 records per second using latency, throughput, and classification performance as evaluation metrics. The results show that the proposed pipeline provides near real-time detection at input rates of up to 200 records per second, with most processing latency remaining below the 5-second Service Level Agreement (SLA) after initialization. In addition, the Random Forest model consistently achieves over 99% classification accuracy across all test scenarios. Performance degradation at 400 records per second is mainly caused by Spark Structured Streaming overhead and data queue accumulation. These findings demonstrate that the proposed architecture is suitable for low-to medium-load DDoS detection and provides a foundation for more scalable stream processing systems.

Keywords: Apache Kafka, Spark Structured Streaming, Random Forest, DDoS, Distributed Stream Processing

KATA PENGANTAR

Dengan memanjatkan puji syukur kepada Tuhan Yang Maha Esa yang telah memberikan rahmat dan karunia-Nya, sehingga tugas akhir ini dapat terselasaikan dengan baik, yang berjudul "PEMANFAATAN APACHE KAFKA DAN SPARK STRUCTURED STREAMING DALAM MONITORING JARINGAN: STUDI KASUS DETEKSI SERANGAN DDOS".

Adapun tujuan dari penulisan tugas akhir ini adalah sebagai bentuk salah satu persyaratan untuk menyelesaikan pendidikan guna mendapatkan gelar Sarjana Strata Satu Program Studi Teknik Informatika, Universitas Mikroskil Medan.

Dalam penyusunan tugas akhir ini, penulis menyampaikan rasa terima kasih kepada:

1. Bapak Dr. Kristian Telaumbanua, S.T., M.T., selaku Dosen Pembimbing I.
2. Bapak Apriyanto Halim, S.Kom., M.Kom., selaku Dosen Pembimbing II.
3. Bapak Hardy, S.Kom., M.Sc., Ph.D., selaku Rektor Universitas Mikroskil Medan.
4. Bapak Sunaryo Winardi, S.Kom., M.T., selaku Dekan Fakultas Informatika Universitas Mikroskil Medan.
5. Bapak Carles Juliandy, S.Kom., M.Kom., selaku Ketua Program Studi S-1 Teknik Informatika Fakultas Informatika Universitas Mikroskil Medan.
6. Bapak dan Ibu Dosen yang telah mendidik dan membimbing penulis selama menempuh pendidikan di Universitas Mikroskil Medan.
7. Keluarga penulis, khususnya orang tua, yang senantiasa memberi dukung moral dan materiil sepanjang awal masa studi hingga saat ini, sehingga penulis dapat menyelesaikan tugas akhir ini.

Penulis menyadari, bahwa penyusunan tugas akhir ini tidak luput dari keterbatasan dan kekurangan. Oleh karena itu, penulis sangat berharap atas kritik dan saran yang bersifat membangun demi penyempurnaan di masa mendatang. Akhir kata, semoga tugas akhir ini dapat memberikan manfaat bagi pembaca dan perkembangan ilmu pengetahuan.

Medan, 24 Juni 2026

Penulis,

FEBRIANTO KUDADIRI

DAFTAR ISI

Abstrak	i
KATA PENGANTAR	ii
DAFTAR ISI	iii
DAFTAR GAMBAR	v
DAFTAR TABEL	ix
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan	2
1.4 Manfaat	2
1.5 Ruang Lingkup	3
BAB II KAJIAN LITERATUR	4
2.1 Keamanan Jaringan Komputer	4
2.2 Intrusion Detection System.....	5
2.3 Distributed Denial of Service	6
2.4 CIC-DDoS 2019	7
2.5 Algoritma Random Forest	8
2.6 Pipeline Data Streaming	13
2.7 Apache Kafka	15
2.8 Apache Avro	16
2.9 Logging.....	17
2.10 Spark Structured Streaming	18
2.11 Apache Parquet	19
2.12 Pemilihan Fitur Model.....	19
2.13 Teknik Penanganan data	21
2.14 FlowChart	24
2.15 Chevron FlowChart	25
2.16 Evaluasi Performa Sistem.....	25
2.17 Penelitian Terdahulu	31
BAB III TAHAPAN PELAKSANAAN	33
3.1 Analisis Masalah.....	33
3.2 Persiapan Data	34
3.3 Pengembangan Model Random Forest	35

3.4 Analisis Proses	48
3.4.1 Analisis Proses Kafka Producer.....	48
3.4.2 Analisis Proses Spark Structured Streaming	53
3.5 Kerangka Pipeline Data Streaming	57
3.6 Skenario Pengujian	64
BAB IV HASIL DAN PENGUJIAN	70
4.1 Hasil.....	70
4.1.1 Persiapan Data	70
4.1.2 Program Random Forest	73
4.1.3 Implementasi Kafka Producer	76
4.1.4 Implementasi Spark Structured Streaming	80
4.2 Pengujian	83
4.2.1 Performa Kafka Producer	84
4.2.2 Performa Spark Structured Streaming	92
4.2.3 Evaluasi Kinerja Random Forest	97
4.2.4 Performa Pipeline Data Streaming	101
4.2.5 Rangkuman Hasil Pengujian.....	111
BAB V PENUTUP	115
DAFTAR PUSTAKA.....	117
DAFTAR RIWAYAT HIDUP	120

DAFTAR GAMBAR

Gambar 2. 1 CIA Triad dalam Keamanan Informasi [11]	4
Gambar 2. 2 Klasifikasi sistem deteksi intrusi menurut metodenya	5
Gambar 2. 3 Mekanisme kerja algoritma Random Forest [17]	10
Gambar 2. 4 Arsitektur data pipeline secara umum [20, 21, 22]	14
Gambar 2. 5 Penggalan struktur skema avro	17
Gambar 2. 6 Simbol-simbol <i>flowchart</i> [36].....	24
Gambar 3. 1 File-file dalam dataset.....	34
Gambar 3. 2 Flowchart Pengembangan Model	35
Gambar 3. 3 Hasil Inisialisasi SparkSession	35
Gambar 3. 4 Contoh Output Preprocessing Data.....	36
Gambar 3. 5 Contoh Output Feature Engineering	36
Gambar 3. 6 Tampilan dataset setelah diurutkan menurut flow duration.....	38
Gambar 3. 7 Visual split data dengan fitur Flow Duration ($t = 13$).....	38
Gambar 3. 8 Visual split data dengan fitur Flow Duration ($t = 504$).....	39
Gambar 3. 9 Visual split data dengan fitur Flow Duration ($t = 3235.5$).....	39
Gambar 3. 10 Tampilan dataset setelah diurutkan menurut Flow Packets/s	40
Gambar 3. 11 Sketsa awal pohon pertama.....	41
Gambar 3. 12 Data yang digunakan berdasarkan ketentuan threshold fitur	41
Gambar 3. 13 Visual split data dengan fitur Flow Packets/s ($t = 40$)	41
Gambar 3. 14 Visual split data dengan fitur Flow Packets/s ($t = 521$)	42
Gambar 3. 15 Visual split data dengan fitur Flow Packets/s ($t = 1500$).....	42
Gambar 3. 16 Tampilan Pohon pertama secara lengkap	43
Gambar 3. 17 Visualisasi Decision Tree 2 sampai 4	44
Gambar 3. 18 Ilustrasi proses prediksi data	44
Gambar 3. 19 Contoh Output Hasil Prediksi Model.....	45
Gambar 3. 20 Multiclass confusion matrix model.....	45
Gambar 3. 21 Binary Confusion Matrix setiap Label.....	46
Gambar 3. 22 Hasil Akurasi Model	46
Gambar 3. 23 Flowchart Kafka Producer	48
Gambar 3. 24 Pemetaan Kolom dan Contoh Gambaran Input	49
Gambar 3. 25 Contoh Gambaran Hasil Split Data	49
Gambar 3. 26 Contoh Gambaran Proses Avro Record Mapping	49
Gambar 3. 27 Contoh Producer Record.....	50

Gambar 3. 28 Contoh Byte Array Hasil Serialisasi	51
Gambar 3. 29 Struktur Pencatatan Log.....	52
Gambar 3. 30 Flowchart Spark Structured Streaming	53
Gambar 3. 31 Proses Identifikasi Penyimpanan Log Kafka	59
Gambar 3. 32 Tampilan Terminal Server Kafka Berjalan.....	60
Gambar 3. 33 Detail Topik	60
Gambar 3. 34 Diagram Alur Kerja Pipeline Yang Dikembangkan	61
Gambar 4. 1 Hasil Pemilihan Data	70
Gambar 4. 2 Hasil Rename Attribut Kolom.....	71
Gambar 4. 3 Banyak Nilai Null dan Infinity Number yang Terdeteksi.....	71
Gambar 4. 4 Hasil Pemeriksaan Nilai Null dan Infinity Setelah Data Cleaning	71
Gambar 4. 5 Dataset Pelatihan Program Random Forest	72
Gambar 4. 6 Hasil Pengacakan Data	72
Gambar 4. 7 Dataset Simulasi Pipeline Streaming	73
Gambar 4. 8 Hasil Pembacaan Data Latih.....	73
Gambar 4. 9 Folder Hasil Penyimpanan Program Random Forest	73
Gambar 4. 10 Struktur Decision Tree Pertama.....	74
Gambar 4. 11 Visual Tree pada Decision Tree Pertama	75
Gambar 4. 12 Nilai-Nilai Features.....	75
Gambar 4. 13 Alur Prediksi Berdasarkan Pohon 1	76
Gambar 4. 14 Persentase Prediksi dan Voting Model	76
Gambar 4. 15 Hasil Pemetaan Header	77
Gambar 4. 16 Hasil Pembacaan Baris Pertama Data.....	77
Gambar 4. 17 Hasil Split Data	77
Gambar 4. 18 Objek Avro Pra-Build	78
Gambar 4. 19 Objek Avro Sesudah Build	78
Gambar 4. 20 Hasil Pembentukan ProducerRecord	79
Gambar 4. 21 Hasil Serialisasi value	79
Gambar 4. 22 Hasil Pencatatan Log	80
Gambar 4. 23 Hasil Pembacaan Data Dari Topic	80
Gambar 4. 24 Hasil Proses Deserialisasi dan seleksi kolom timestamp	81
Gambar 4. 25 Hasil Ekstraksi Kolom Data	81
Gambar 4. 26 Hasil Penambahan Kolom kafkaIngestTime	82
Gambar 4. 27 Hasil Data Feature Engineering	82
Gambar 4. 28 Data Hasil Prediksi Model	82

Gambar 4. 29 Data Hasil Deteksi DDoS	83
Gambar 4. 30 Hasil Penyimpanan Output Spark	83
Gambar 4. 31 Hasil Pengujian Latency (Percobaan 1)	84
Gambar 4. 32 Pseudocode Diagram Perbandingan Persentil Latency	84
Gambar 4. 33 Perbandingan Persentil Latency pada Percobaan 1	85
Gambar 4. 34 Pseudocode Diagram Perbandingan Latency Tiap Percobaan	85
Gambar 4. 35 Perbandingan P95 Latency ketiga Percobaan	86
Gambar 4. 36 Data Hasil Pengujian Latency (Percobaan 2)	86
Gambar 4. 37 Perbandingan Persentil Latency (Percobaan 2)	87
Gambar 4. 38 Perbandingan P95 Latency ketiga Percobaan	87
Gambar 4. 39 Data Hasil Pengujian Latency (Percobaan 3)	88
Gambar 4. 40 Perbandingan Persentil Latency (Percobaan 3)	88
Gambar 4. 41 Perbandingan P95 Latency ketiga Percobaan	88
Gambar 4. 42 Pseudocode Diagram Throughput	90
Gambar 4. 43 Grafik Pengujian Throughput (Percobaan 1)	90
Gambar 4. 44 Grafik Pengujian Throughput (Percobaan 2)	91
Gambar 4. 45 Grafik Pengujian Throughput (Percobaan 3)	91
Gambar 4. 46 Hasil Pengujian Spark (Percobaan 1)	93
Gambar 4. 47 Pseudocode Diagram Latency dan Throughput Spark	93
Gambar 4. 48 Grafik Performa Spark (Percobaan 1)	93
Gambar 4. 49 Hasil Pengujian Spark (Percobaan 2)	94
Gambar 4. 50 Grafik Performa Spark (Percobaan 2)	94
Gambar 4. 51 Hasil Pengujian Spark (Percobaan 3)	95
Gambar 4. 52 Grafik Performa Spark (Percobaan 3)	95
Gambar 4. 53 Hasil Pengujian Spark (Percobaan 1)	96
Gambar 4. 54 Grafik Performa Spark (Percobaan 1)	96
Gambar 4. 55 Confusion Matrix Random Forest (Percobaan 1)	98
Gambar 4. 56 Classification Report Random Forest (Percobaan 1)	98
Gambar 4. 57 Distribusi Klasifikasi Prediksi Random Forest (percobaan 1)	98
Gambar 4. 58 Confusion Matrix Random Forest (Percobaan 3)	99
Gambar 4. 59 Classification Report Random Forest (Percobaan 3)	99
Gambar 4. 60 Distribusi Klasifikasi Prediksi Random Forest (percobaan 3)	99
Gambar 4. 61 Confusion Matrix Random Forest (Percobaan 2)	100
Gambar 4. 62 Classification Report Random Forest (Percobaan 2)	100
Gambar 4. 63 Distribusi Klasifikasi Prediksi Random Forest (percobaan 2)	100

Gambar 4. 64 Hasil Pengujian Latency Pipeline (percobaan 2)	101
Gambar 4. 65 Perbandingan Persentil Latensi pipeline (percobaan 2)	101
Gambar 4. 66 Perbandingan P95 Latency Pipeline pada Tiga Percobaan	102
Gambar 4. 67 Hasil Pengujian Latency Pipeline (Percobaan 1)	103
Gambar 4. 68 Perbandingan Persentil Latensi pipeline (percobaan 1)	103
Gambar 4. 69 Hasil Pengujian Latency Pipeline (Percobaan 3)	103
Gambar 4. 70 Perbandingan Persentil Latensi pipeline (percobaan 3)	104
Gambar 4. 71 Hasil Pengujian Latency Pipeline (Percobaan 2)	104
Gambar 4. 72 Perbandingan Persentil Latensi pipeline (percobaan 2)	105
Gambar 4. 73 Perbandingan P95 Latency Pipeline pada Tiga Percobaan	105
Gambar 4. 74 Hasil Pengujian Throughput Pipeline (Percobaan 2)	106
Gambar 4. 75 Diagram Data yang Diproses Pipeline Tiap Interval	106
Gambar 4. 76 Throughput Aktual Pipeline Tiap Interval Waktu	107
Gambar 4. 77 Hasil Pengujian Throughput Pipeline (Percobaan 3)	107
Gambar 4. 78 Diagram Data yang Diproses Pipeline Tiap Interval	108
Gambar 4. 79 Throughput Aktual Pipeline Tiap Interval Waktu	108
Gambar 4. 80 Hasil Pengujian Throughput Pipeline (Percobaan 2)	109
Gambar 4. 81 Diagram Data yang Diproses Pipeline Tiap Interval	109
Gambar 4. 82 Throughput Aktual Pipeline Tiap Interval Waktu	110
Gambar 4. 83 Validasi Pengujian Performa Skenario 100 record/s	111
Gambar 4. 84 Validasi Pengujian Performa Skenario 200 record/s	112
Gambar 4. 85 Validasi Pengujian Performa Skenario 400 record/s	112

DAFTAR TABEL

Tabel 2. 1 Identifikasi Sebagian Atribut Dataset [13]	7
Tabel 2. 2 Pemilihan Fitur [13], [14]	20
Tabel 2. 3 Kategorisasi <i>Latency</i> komponen Kafka [40]	29
Tabel 2. 4 Kategorisasi <i>latency Spark dan pipeline E2E near real-time</i> [40], [41].	30
Tabel 2. 5 Kategorisasi <i>Throughput</i> [3], [8], [19]	31
Tabel 2. 6 Penelitian Terdahulu	31
Tabel 3. 1 Contoh Dataset Model	37
Tabel 3. 2 Data training pohon 1 berdasarkan bootstrap sampling	38
Tabel 3. 3 Dataset test	44
Tabel 3. 4 Hasil Perhitungan ASCII	51
Tabel 3. 5 Skema Kolom Hasil Pembacaan Data Dari Topik	53
Tabel 3. 6 Skema Kolom Hasil Deserialisasi Avro	54
Tabel 3. 7 Skema Kolom Hasil Ekstaksi Kolom	55
Tabel 3. 8 Struktur Data Setelah Penambahan Kolom	55
Tabel 3. 9 Skema Kolom Dataset Model	56
Tabel 3. 10 Skema Kolom Hasil Prediksi Model	56
Tabel 3. 11 Skema Kolom Hasil Deteksi DDoS	57
Tabel 3. 12 Aturan dan Tujuan Pengujian	65
Tabel 3. 13 Metrik Yang Digunakan	65
Tabel 3. 14 Struktur Data Laporan StreamingQueryListener	67
Tabel 3. 15 Struktur Data Pengujian Kafka Producer	68
Tabel 3. 16 Struktur Data Pengujian Pipeline E2E: Latency	68
Tabel 3. 17 Struktur Data Pengujian Pipeline E2E: Throughput	69