

Makalah Tugas 1: Perancangan Arsitektur Big Data

Arsitektur Big Data untuk Analisis Sentimen Twitter Menggunakan Layanan AWS

Dosen Pengampu: Adam Sekti Aji, S.Kom., M.Kom



Disusun oleh:

Lathif Ramadhan (5231811022)

PROGRAM STUDI SAINS DATA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS TEKNOLOGI YOGYAKARTA

YOGYAKARTA

2024

Daftar Isi

Daftar Isi	II
BAB 1: Pendahuluan	1
1.1. Latar Belakang	1
1.2. Tujuan	1
1.3. Manfaat	1
1. Memahami Konsep Arsitektur Big Data	2
2. Mempelajari Integrasi Layanan AWS.....	2
3. Mendapatkan Insights dari Data Twitter.....	2
4. Meningkatkan Efisiensi dan Skalabilitas.....	2
5. Membuka Peluang untuk Inovasi	2
BAB 2: Arsitektur Sistem	3
2.1. Diagram Arsitektur.....	3
2.2. Komponen Utama Arsitektur	3
2.2.1. Data Ingestion (Pengumpulan Data).....	3
2.2.2. Data Storage (Penyimpanan Data).....	4
2.2.3. Data Processing (Pengolahan Data).....	4
2.2.4. Data Analysis (Analisis Data).....	5
2.2.5. Data Consumption (Konsumsi Data).....	5
BAB 3: Flowchart Proses Analisis Sentimen	6
3.1. Diagram Flowchart	6
3.2. Alur Kerja Sistem.....	6
3.2.1. Pengumpulan Data	6
3.2.2. Pengiriman Data.....	7
3.2.3. Pemrosesan Data.....	7
3.2.4. Analisis Sentimen.....	7
3.2.5. Query Data	8
3.2.6. Visualisasi Data.....	8
BAB 4: Penjelasan Komponen.....	9
4.1. Twitter API.....	9
• Fungsi Utama	9
• Keunggulan	9

• Contoh Penggunaan.....	9
4.2. Amazon Kinesis Data Firehose	9
• Fungsi Utama	9
• Keunggulan	9
• Contoh Penggunaan.....	9
4.3. Amazon S3	10
• Fungsi Utama	10
• Keunggulan	10
• Contoh Penggunaan.....	10
4.4. AWS Lambda	10
• Fungsi Utama	10
• Keunggulan	10
• Contoh Penggunaan.....	10
4.5. Amazon Comprehend.....	11
• Fungsi Utama	11
• Keunggulan	11
• Contoh Penggunaan.....	11
4.6. Amazon Athena.....	11
• Fungsi Utama	11
• Keunggulan	11
• Contoh Penggunaan.....	11
4.7. Amazon QuickSight.....	11
• Fungsi Utama	11
• Keunggulan	12
• Contoh Penggunaan.....	12
BAB 5: Kesimpulan	13
5.1. Skalabilitas dan Fleksibilitas.....	13
5.2. Real-Time Processing	13
5.3. Integrasi yang Mudah	13
5.4. Manfaat bagi Bisnis	13
5.5. Potensi Pengembangan	14
5.6. Kesimpulan Akhir	14

BAB 6: Sumber dan Referensi.....	15
---	-----------

BAB 1: Pendahuluan

1.1. Latar Belakang

Di era digital seperti sekarang, media sosial telah menjadi salah satu platform utama bagi masyarakat untuk mengekspresikan pendapat, emosi, dan reaksi terhadap berbagai topik, mulai dari produk, layanan, hingga isu sosial dan politik. Twitter, sebagai salah satu platform media sosial terbesar, menghasilkan jutaan tweet setiap harinya. Data ini tidak hanya berisi informasi tekstual, tetapi juga mencerminkan sentimen atau perasaan pengguna terhadap suatu topik tertentu.

Analisis sentimen, atau sentiment analysis, adalah proses mengidentifikasi dan mengkategorikan opini yang terkandung dalam teks, seperti tweet, untuk menentukan apakah opini tersebut bersifat positif, negatif, atau netral. Dengan menganalisis sentimen dari tweet, perusahaan atau organisasi dapat memahami bagaimana publik merespons produk, layanan, atau kampanye mereka. Informasi ini sangat berharga untuk pengambilan keputusan strategis, seperti meningkatkan kualitas produk, merespons keluhan pelanggan, atau mengidentifikasi tren pasar.

Namun, mengolah data Twitter dalam skala besar bukanlah hal yang mudah. Data yang dihasilkan oleh Twitter bersifat real-time, tidak terstruktur, dan memiliki volume yang sangat besar. Di sinilah teknologi Big Data dan layanan cloud computing seperti Amazon Web Services (AWS) memainkan peran penting. AWS menyediakan berbagai layanan yang dapat diintegrasikan untuk membangun arsitektur Big Data yang scalable, efisien, dan mampu menangani data dalam skala besar secara real-time.

1.2. Tujuan

Makalah ini bertujuan untuk menjelaskan bagaimana merancang dan mengimplementasikan arsitektur Big Data yang menggunakan layanan AWS untuk melakukan analisis sentimen pada data Twitter. Arsitektur ini dirancang untuk menangani alur kerja mulai dari pengumpulan data, penyimpanan, pemrosesan, hingga analisis dan visualisasi data. Dengan memanfaatkan layanan AWS seperti Amazon Kinesis, Amazon S3, AWS Lambda, dan Amazon Comprehend, sistem ini dapat mengolah data Twitter secara efisien dan menghasilkan insights yang berguna bagi pengguna bisnis.

Selain itu, makalah ini juga bertujuan untuk memberikan pemahaman mendalam tentang bagaimana setiap komponen dalam arsitektur ini bekerja, mulai dari bagaimana data dikumpulkan dari Twitter, disimpan di Amazon S3, diproses menggunakan AWS Lambda, hingga dianalisis menggunakan Amazon Comprehend. Dengan memahami arsitektur ini, pembaca diharapkan dapat mengaplikasikan konsep yang sama dalam proyek-proyek Big Data lainnya.

1.3. Manfaat

Makalah ini memberikan beberapa manfaat penting, baik dari segi teoritis maupun praktis:

1. Memahami Konsep Arsitektur Big Data

Pembaca akan mendapatkan pemahaman mendalam tentang bagaimana merancang arsitektur Big Data yang scalable dan efisien menggunakan layanan AWS. Arsitektur ini tidak hanya terbatas pada analisis sentimen Twitter, tetapi juga dapat diaplikasikan pada berbagai kasus penggunaan lainnya.

2. Mempelajari Integrasi Layanan AWS

Makalah ini menjelaskan bagaimana berbagai layanan AWS seperti Amazon Kinesis, Amazon S3, AWS Lambda, dan Amazon Comprehend dapat diintegrasikan untuk membangun sistem yang komprehensif. Pembaca akan belajar bagaimana setiap layanan ini saling melengkapi dan bekerja sama dalam satu arsitektur.

3. Mendapatkan Insights dari Data Twitter

Dengan menerapkan arsitektur ini, perusahaan atau organisasi dapat mengumpulkan dan menganalisis data Twitter secara real-time. Hasil analisis sentimen ini dapat digunakan untuk memahami opini publik, mengidentifikasi tren, dan mengambil keputusan bisnis yang lebih baik.

4. Meningkatkan Efisiensi dan Skalabilitas

Arsitektur yang dibangun menggunakan layanan AWS memungkinkan sistem untuk menangani data dalam skala besar tanpa perlu khawatir tentang infrastruktur fisik. Selain itu, sistem ini dapat dengan mudah di-scale up atau down sesuai kebutuhan.

5. Membuka Peluang untuk Inovasi

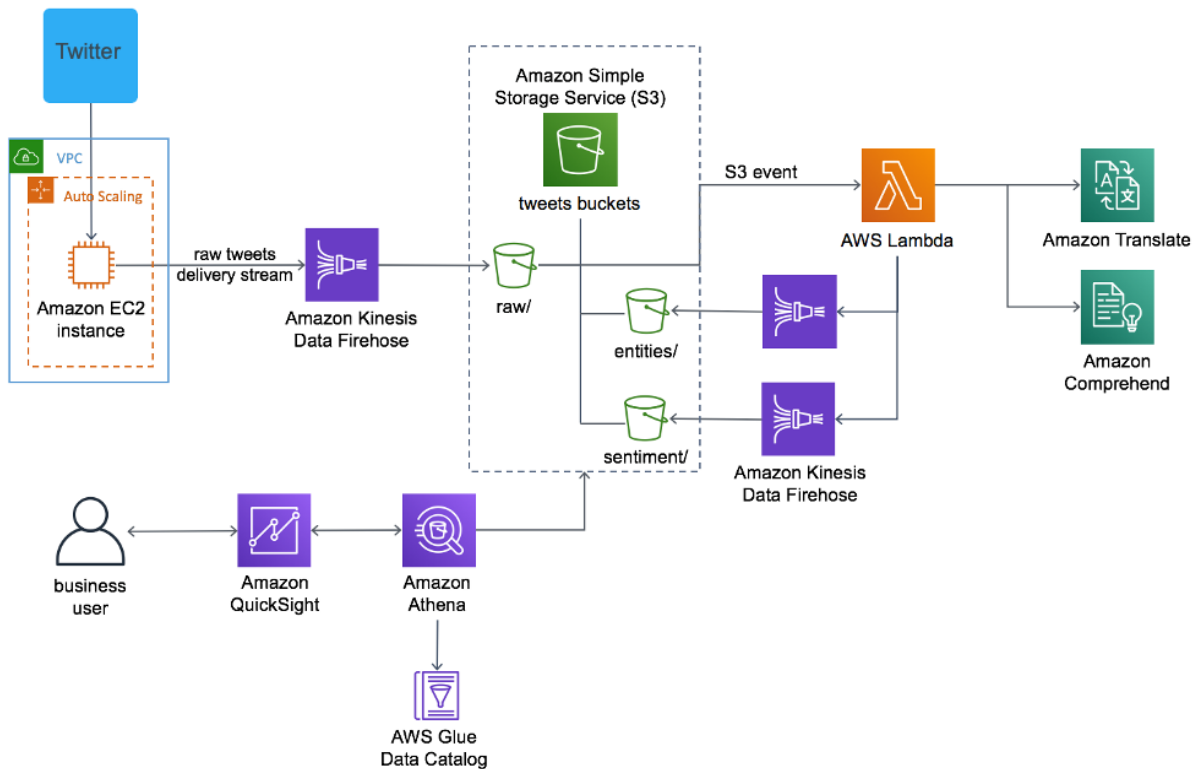
Dengan memahami cara kerja arsitektur ini, pembaca dapat mengembangkan solusi-solusi inovatif lainnya yang memanfaatkan Big Data dan layanan cloud computing. Misalnya, sistem ini dapat dikembangkan lebih lanjut untuk analisis data dari platform media sosial lainnya atau untuk kasus penggunaan yang lebih kompleks.

Dengan demikian, makalah ini tidak hanya memberikan panduan teknis tentang bagaimana membangun arsitektur Big Data untuk analisis sentimen Twitter, tetapi juga memberikan wawasan tentang bagaimana teknologi ini dapat memberikan nilai tambah bagi bisnis dan organisasi.

BAB 2: Arsitektur Sistem

2.1. Diagram Arsitektur

Berikut adalah diagram arsitektur sistem berdasarkan file yang diberikan:



2.2. Komponen Utama Arsitektur

Berdasarkan diagram dibawah, arsitektur sistem untuk analisis sentimen Twitter menggunakan layanan AWS terdiri dari beberapa komponen utama yang saling terintegrasi. Setiap komponen memiliki peran dan fungsi khusus dalam mengelola alur data, mulai dari pengumpulan, penyimpanan, pemrosesan, analisis, hingga visualisasi data. Berikut adalah penjelasan rinci tentang setiap komponen:

2.2.1. Data Ingestion (Pengumpulan Data)

Pengumpulan data adalah tahap pertama dalam arsitektur Big Data. Pada tahap ini, data tweet dikumpulkan dari Twitter dan dikirim ke sistem penyimpanan untuk diproses lebih lanjut.

- **Twitter API**

Twitter menyediakan API (Application Programming Interface) yang memungkinkan pengembang untuk mengakses data tweet secara real-time. API ini dapat dikonfigurasi untuk mengumpulkan tweet berdasarkan kata kunci, hashtag, lokasi, atau parameter lainnya. Data yang dikumpulkan mencakup teks tweet, informasi pengguna, waktu posting, dan metadata lainnya.

Contohnya, jika perusahaan ingin menganalisis sentimen tentang produk mereka, mereka dapat mengumpulkan tweet yang mengandung nama produk atau hashtag terkait.

- **Amazon Kinesis Data Firehose**

Setelah data tweet dikumpulkan melalui Twitter API, data tersebut perlu dikirim ke sistem penyimpanan dengan cepat dan efisien. Amazon Kinesis Data Firehose adalah layanan yang dirancang untuk menangani aliran data real-time dalam skala besar. Kinesis Firehose menerima data dari Twitter API, memprosesnya secara minimal (jika diperlukan), dan mengirimkannya ke Amazon S3. Layanan ini juga dapat mengompres atau mengenkripsi data sebelum menyimpannya, sehingga memastikan keamanan dan efisiensi penyimpanan.

2.2.2. Data Storage (Penyimpanan Data)

Setelah data dikumpulkan, data tersebut perlu disimpan dalam sistem yang scalable dan aman. AWS menyediakan layanan penyimpanan yang cocok untuk menangani data dalam skala besar.

- **Amazon S3 (Simple Storage Service)**

Amazon S3 adalah layanan penyimpanan objek yang sangat scalable dan andal. Data tweet yang dikirim melalui Kinesis Firehose disimpan dalam bucket S3. Bucket ini diorganisir dengan struktur folder seperti raw/entities/ dan sentiment/ untuk memudahkan pengelolaan data.

Misalnya, data mentah dari Twitter disimpan di folder raw/, sementara hasil analisis sentimen disimpan di folder sentiment/. Struktur ini memudahkan tim data untuk mengakses dan memproses data sesuai kebutuhan.

- **AWS Glue Data Catalog**

AWS Glue Data Catalog adalah layanan yang digunakan untuk menyimpan metadata tentang data yang disimpan di S3. Metadata ini mencakup informasi seperti struktur data, lokasi file, dan format data.

Dengan menggunakan Glue Data Catalog, tim data dapat dengan mudah melakukan query pada data yang disimpan di S3 menggunakan Amazon Athena, tanpa perlu memindahkan atau mengubah format data.

2.2.3. Data Processing (Pengolahan Data)

Setelah data disimpan, langkah selanjutnya adalah memproses data untuk mempersiapkannya ke tahap analisis. Pemrosesan data melibatkan pembersihan, transformasi, dan analisis awal.

- **AWS Lambda:**

AWS Lambda adalah layanan komputasi tanpa server yang memungkinkan pengembang untuk menjalankan kode sebagai respons terhadap event tertentu. Dalam arsitektur ini, Lambda dipicu oleh event S3, yaitu ketika data baru masuk ke bucket S3.

Fungsi Lambda dapat digunakan untuk memproses data tweet, seperti membersihkan data dari karakter yang tidak diperlukan, memfilter tweet yang tidak relevan, atau memformat data agar siap untuk analisis.

- **Amazon Translate:**

Twitter adalah platform global, sehingga banyak tweet yang mungkin ditulis dalam berbagai bahasa. Amazon Translate adalah layanan terjemahan otomatis yang dapat menerjemahkan tweet ke bahasa yang diinginkan, misalnya bahasa Inggris.

Dengan menerjemahkan tweet, analisis sentimen dapat dilakukan secara lebih akurat,

terutama jika model analisis sentimen yang digunakan hanya mendukung bahasa tertentu.

- **Amazon Comprehend:**

Amazon Comprehend adalah layanan Natural Language Processing (NLP) yang menggunakan Machine Learning untuk menganalisis teks. Layanan ini dapat mengidentifikasi sentimen (positif, negatif, atau netral), entitas (seperti nama orang, organisasi, atau lokasi), dan topik yang dibahas dalam tweet. Misalnya, Comprehend dapat menganalisis tweet tentang suatu produk dan menentukan apakah tweet tersebut mengandung sentimen positif, negatif, atau netral.

2.2.4. Data Analysis (Analisis Data)

Setelah data diproses, langkah selanjutnya adalah menganalisis data untuk menghasilkan insights yang berguna.

- **Amazon Athena:**

Amazon Athena adalah layanan query yang memungkinkan pengguna untuk melakukan query SQL pada data yang disimpan di S3. Athena tidak memerlukan infrastruktur khusus, karena query dijalankan langsung pada data di S3. Misalnya, tim data dapat menggunakan Athena untuk mencari tweet yang mengandung sentimen negatif atau untuk menghitung jumlah tweet positif tentang suatu topik.

- **Amazon QuickSight:**

Amazon QuickSight adalah layanan visualisasi data yang memungkinkan pengguna untuk membuat dashboard interaktif. Hasil analisis dari Amazon Athena dapat divisualisasikan menggunakan QuickSight, sehingga pengguna bisnis dapat dengan mudah memahami tren dan pola dalam data. Contohnya, QuickSight dapat menampilkan grafik yang menunjukkan persentase tweet positif, negatif, dan netral tentang suatu produk.

2.2.5. Data Consumption (Konsumsi Data)

Tahap terakhir dalam arsitektur ini adalah konsumsi data, di mana hasil analisis digunakan oleh pengguna bisnis untuk pengambilan keputusan.

- **Business User:**

Pengguna bisnis, seperti manajer produk atau tim pemasaran, mengakses hasil analisis melalui dashboard Amazon QuickSight. Dashboard ini menyajikan informasi dalam bentuk visual yang mudah dipahami, seperti grafik, tabel, dan laporan. Dengan informasi ini, pengguna bisnis dapat mengambil keputusan yang lebih baik, seperti meningkatkan kualitas produk, merespons keluhan pelanggan, atau mengidentifikasi peluang pasar.

BAB 3: Flowchart Proses Analisis Sentimen

3.1. Diagram Flowchart

Berikut adalah diagram flowchart yang menjelaskan alur kerja sistem:



3.2. Alur Kerja Sistem

Flowchart ini menjelaskan alur kerja sistem analisis sentimen Twitter menggunakan layanan AWS, mulai dari pengumpulan data hingga visualisasi hasil analisis. Setiap tahap dalam flowchart ini dirancang untuk memastikan bahwa data diproses secara efisien dan menghasilkan insights yang berguna bagi pengguna bisnis. Berikut adalah penjelasan rinci tentang setiap langkah dalam alur kerja sistem:

3.2.1. Pengumpulan Data

Tahap pertama dalam alur kerja ini adalah pengumpulan data tweet dari Twitter.

- **Twitter API:**

Twitter menyediakan API yang memungkinkan pengembang untuk mengakses data tweet secara real-time. API ini dapat dikonfigurasi untuk mengumpulkan tweet berdasarkan kata kunci, hashtag, lokasi, atau parameter lainnya.

Misalnya, jika perusahaan ingin menganalisis sentimen tentang produk mereka, mereka dapat mengumpulkan tweet yang mengandung nama produk atau hashtag terkait. Data yang dikumpulkan mencakup teks tweet, informasi pengguna, waktu posting, dan metadata lainnya.

- **Keunggulan:**

Twitter API sangat fleksibel dan dapat disesuaikan dengan kebutuhan analisis. Selain itu, API ini mendukung streaming data, sehingga data dapat dikumpulkan secara real-time.

3.2.2. Pengiriman Data

Setelah data tweet dikumpulkan, data tersebut perlu dikirim ke sistem penyimpanan untuk diproses lebih lanjut.

- **Amazon Kinesis Data Firehose:**
Kinesis Firehose adalah layanan yang dirancang untuk menangani aliran data real-time dalam skala besar. Layanan ini menerima data dari Twitter API, memprosesnya secara minimal (jika diperlukan), dan mengirimkannya ke Amazon S3. Kinesis Firehose juga dapat mengompres atau mengenkripsi data sebelum menyimpannya, sehingga memastikan keamanan dan efisiensi penyimpanan.
- **Keunggulan:**
Kinesis Firehose sangat scalable dan dapat menangani jutaan data per detik. Layanan ini juga terintegrasi dengan baik dengan layanan AWS lainnya, seperti S3 dan Lambda.

3.2.3. Pemrosesan Data

Setelah data disimpan di Amazon S3, langkah selanjutnya adalah memproses data untuk mempersiapkannya ke tahap analisis.

- **AWS Lambda:**
AWS Lambda adalah layanan komputasi tanpa server yang memungkinkan pengembang untuk menjalankan kode sebagai respons terhadap event tertentu. Dalam arsitektur ini, Lambda dipicu oleh event S3, yaitu ketika data baru masuk ke bucket S3. Fungsi Lambda dapat digunakan untuk memproses data tweet, seperti membersihkan data dari karakter yang tidak diperlukan, memfilter tweet yang tidak relevan, atau memformat data agar siap untuk analisis.
- **Keunggulan:**
Lambda adalah layanan tanpa server, artinya pengembang tidak perlu mengelola infrastruktur. Layanan ini juga sangat scalable dan dapat menangani ribuan event secara bersamaan.

3.2.4. Analisis Sentimen

Setelah data diproses, langkah selanjutnya adalah menganalisis sentimen dari tweet.

- **Amazon Comprehend:**
Amazon Comprehend adalah layanan Natural Language Processing (NLP) yang menggunakan Machine Learning untuk menganalisis teks. Layanan ini dapat mengidentifikasi sentimen (positif, negatif, atau netral), entitas (seperti nama orang, organisasi, atau lokasi), dan topik yang dibahas dalam tweet. Misalnya, Comprehend dapat menganalisis tweet tentang suatu produk dan menentukan apakah tweet tersebut mengandung sentimen positif, negatif, atau netral.
- **Keunggulan:**
Comprehend sangat akurat dan dapat menangani data dalam skala besar. Layanan ini juga terintegrasi dengan baik dengan layanan AWS lainnya, seperti S3 dan Lambda.

3.2.5. Query Data

Setelah data dianalisis, langkah selanjutnya adalah melakukan query pada data untuk menghasilkan insights yang berguna.

- **Amazon Athena:**

Amazon Athena adalah layanan query yang memungkinkan pengguna untuk melakukan query SQL pada data yang disimpan di S3. Athena tidak memerlukan infrastruktur khusus, karena query dijalankan langsung pada data di S3.

Misalnya, tim data dapat menggunakan Athena untuk mencari tweet yang mengandung sentimen negatif atau untuk menghitung jumlah tweet positif tentang suatu topik.

- **Keunggulan:**

Athena sangat mudah digunakan dan tidak memerlukan setup infrastruktur. Layanan ini juga sangat cepat dan dapat menangani data dalam skala besar.

3.2.6. Visualisasi Data

Tahap terakhir dalam alur kerja ini adalah visualisasi data, di mana hasil analisis ditampilkan dalam bentuk yang mudah dipahami oleh pengguna bisnis.

- **Amazon QuickSight:**

Amazon QuickSight adalah layanan visualisasi data yang memungkinkan pengguna untuk membuat dashboard interaktif. Hasil analisis dari Amazon Athena dapat divisualisasikan menggunakan QuickSight, sehingga pengguna bisnis dapat dengan mudah memahami tren dan pola dalam data.

Contohnya, QuickSight dapat menampilkan grafik yang menunjukkan persentase tweet positif, negatif, dan netral tentang suatu produk.

- **Keunggulan:**

QuickSight sangat mudah digunakan dan menyediakan berbagai fitur visualisasi yang interaktif. Layanan ini juga terintegrasi dengan baik dengan layanan AWS lainnya, seperti Athena dan S3.

BAB 4: Penjelasan Komponen

4.1. Twitter API

Twitter API adalah antarmuka pemrograman aplikasi yang disediakan oleh Twitter untuk mengakses data tweet secara real-time. API ini memungkinkan pengembang untuk mengambil data dari Twitter berdasarkan parameter tertentu, seperti kata kunci, hashtag, lokasi, atau bahkan ID pengguna.

- **Fungsi Utama:**

Twitter API digunakan untuk mengumpulkan data tweet yang relevan dengan topik atau tujuan analisis. Misalnya, jika perusahaan ingin menganalisis sentimen tentang produk mereka, mereka dapat mengumpulkan tweet yang mengandung nama produk atau hashtag terkait.

API ini juga menyediakan metadata tambahan, seperti waktu posting, lokasi pengguna, dan informasi akun, yang dapat berguna untuk analisis lebih lanjut.

- **Keunggulan:**

Twitter API sangat fleksibel dan dapat dikustomisasi sesuai kebutuhan. Pengembang dapat memilih jenis data yang ingin dikumpulkan, seperti tweet publik, retweet, atau bahkan pesan langsung (DM). Selain itu, API ini mendukung streaming data, sehingga data dapat dikumpulkan secara real-time.

- **Contoh Penggunaan:**

Sebuah perusahaan e-commerce dapat menggunakan Twitter API untuk mengumpulkan tweet yang mengandung nama merek mereka. Data ini kemudian dapat dianalisis untuk memahami bagaimana pelanggan merespons produk atau kampanye pemasaran terbaru.

4.2. Amazon Kinesis Data Firehose

Amazon Kinesis Data Firehose adalah layanan yang dirancang untuk menangani aliran data real-time dalam skala besar. Layanan ini bertindak sebagai jembatan antara sumber data (seperti Twitter API) dan sistem penyimpanan (seperti Amazon S3).

- **Fungsi Utama:**

Kinesis Firehose menerima data dari Twitter API, memprosesnya secara minimal (jika diperlukan), dan mengirimkannya ke Amazon S3. Layanan ini juga dapat mengompres atau mengenkripsi data sebelum menyimpannya, sehingga memastikan keamanan dan efisiensi penyimpanan.

- **Keunggulan:**

Kinesis Firehose sangat scalable dan dapat menangani jutaan data per detik. Layanan ini juga terintegrasi dengan baik dengan layanan AWS lainnya, seperti Lambda dan S3, sehingga memudahkan pengembang dalam membangun arsitektur yang kompleks.

- **Contoh Penggunaan:**

Dalam sistem analisis sentimen, Kinesis Firehose digunakan untuk mengirim data

tweet dari Twitter API ke Amazon S3. Data ini kemudian dapat diproses lebih lanjut menggunakan layanan AWS lainnya.

4.3. Amazon S3

Amazon S3 (Simple Storage Service) adalah layanan penyimpanan objek yang sangat scalable dan aman. S3 menyediakan penyimpanan yang andal untuk berbagai jenis data, termasuk data mentah, data yang telah diproses, dan hasil analisis.

- **Fungsi Utama:**

Data tweet yang dikirim melalui Kinesis Firehose disimpan dalam bucket S3. Bucket ini diorganisir dengan struktur folder seperti raw/entities/ dan sentiment/ untuk memudahkan pengelolaan data.

Misalnya, data mentah dari Twitter disimpan di folder raw/, sementara hasil analisis sentimen disimpan di folder sentiment/.

- **Keunggulan:**

Amazon S3 sangat scalable, artinya dapat menangani data dalam skala besar tanpa perlu khawatir tentang kapasitas penyimpanan. Selain itu, S3 menyediakan fitur keamanan seperti enkripsi data dan kontrol akses, sehingga memastikan data tetap aman.

- **Contoh Penggunaan:**

Dalam sistem analisis sentimen, Amazon S3 digunakan untuk menyimpan data tweet mentah dan hasil analisis sentimen. Data ini kemudian dapat diakses oleh layanan AWS lainnya, seperti Athena dan QuickSight.

4.4. AWS Lambda

AWS Lambda adalah layanan komputasi tanpa server yang memungkinkan pengembang untuk menjalankan kode sebagai respons terhadap event tertentu. Dalam arsitektur ini, Lambda dipicu oleh event S3, yaitu ketika data baru masuk ke bucket S3.

- **Fungsi Utama:**

Fungsi Lambda dapat digunakan untuk memproses data tweet, seperti membersihkan data dari karakter yang tidak diperlukan, memfilter tweet yang tidak relevan, atau memformat data agar siap untuk analisis.

- **Keunggulan:**

Lambda adalah layanan tanpa server, artinya pengembang tidak perlu mengelola infrastruktur. Layanan ini juga sangat scalable dan dapat menangani ribuan event secara bersamaan.

- **Contoh Penggunaan:**

Dalam sistem analisis sentimen, Lambda digunakan untuk memproses data tweet yang baru masuk ke S3. Misalnya, Lambda dapat membersihkan data tweet dari karakter khusus atau memformat data agar siap untuk analisis menggunakan Amazon Comprehend.

4.5. Amazon Comprehend

Amazon Comprehend adalah layanan Natural Language Processing (NLP) yang menggunakan Machine Learning untuk menganalisis teks. Layanan ini dapat mengidentifikasi sentimen, entitas, dan topik yang dibahas dalam teks.

- **Fungsi Utama:**
Comprehend digunakan untuk menganalisis sentimen dari tweet. Layanan ini dapat mengidentifikasi apakah tweet tersebut mengandung sentimen positif, negatif, atau netral. Selain itu, Comprehend juga dapat mengidentifikasi entitas seperti nama orang, organisasi, atau lokasi yang disebutkan dalam tweet.
- **Keunggulan:**
Comprehend sangat akurat dan dapat menangani data dalam skala besar. Layanan ini juga terintegrasi dengan baik dengan layanan AWS lainnya, seperti S3 dan Lambda.
- **Contoh Penggunaan:**
Dalam sistem analisis sentimen, Comprehend digunakan untuk menganalisis sentimen dari tweet yang telah diproses oleh Lambda. Hasil analisis ini kemudian disimpan kembali ke S3 untuk digunakan lebih lanjut.

4.6. Amazon Athena

Amazon Athena adalah layanan query yang memungkinkan pengguna untuk melakukan query SQL pada data yang disimpan di S3. Athena tidak memerlukan infrastruktur khusus, karena query dijalankan langsung pada data di S3.

- **Fungsi Utama:**
Athena digunakan untuk melakukan query pada data yang disimpan di S3. Misalnya, tim data dapat menggunakan Athena untuk mencari tweet yang mengandung sentimen negatif atau untuk menghitung jumlah tweet positif tentang suatu topik.
- **Keunggulan:**
Athena sangat mudah digunakan dan tidak memerlukan setup infrastruktur. Layanan ini juga sangat cepat dan dapat menangani data dalam skala besar.
- **Contoh Penggunaan:**
Dalam sistem analisis sentimen, Athena digunakan untuk melakukan query pada data tweet yang telah dianalisis oleh Comprehend. Hasil query ini kemudian dapat divisualisasikan menggunakan QuickSight.

4.7. Amazon QuickSight

Amazon QuickSight adalah layanan visualisasi data yang memungkinkan pengguna untuk membuat dashboard interaktif. Hasil analisis dari Amazon Athena dapat divisualisasikan menggunakan QuickSight, sehingga pengguna bisnis dapat dengan mudah memahami tren dan pola dalam data.

- **Fungsi Utama:**
QuickSight digunakan untuk menampilkan hasil analisis sentimen dalam bentuk

visual, seperti grafik, tabel, dan laporan. Dashboard ini dapat diakses oleh pengguna bisnis untuk mengambil keputusan yang lebih baik.

- **Keunggulan:**

QuickSight sangat mudah digunakan dan menyediakan berbagai fitur visualisasi yang interaktif. Layanan ini juga terintegrasi dengan baik dengan layanan AWS lainnya, seperti Athena dan S3.

- **Contoh Penggunaan:**

Dalam sistem analisis sentimen, QuickSight digunakan untuk menampilkan grafik yang menunjukkan persentase tweet positif, negatif, dan netral tentang suatu produk. Dashboard ini dapat diakses oleh tim pemasaran untuk memahami respons pelanggan terhadap produk mereka.

BAB 5: Kesimpulan

Dalam makalah ini, kita telah membahas secara mendetail tentang bagaimana membangun arsitektur Big Data untuk analisis sentimen Twitter menggunakan layanan Amazon Web Services (AWS). Arsitektur ini dirancang untuk menangani alur kerja mulai dari pengumpulan data, penyimpanan, pemrosesan, analisis, hingga visualisasi data. Dengan memanfaatkan layanan AWS seperti **Amazon Kinesis**, **Amazon S3**, **AWS Lambda**, dan **Amazon Comprehend**, sistem ini tidak hanya efisien tetapi juga scalable, real-time, dan mudah diintegrasikan dengan berbagai layanan analisis lainnya.

5.1. Skalabilitas dan Fleksibilitas

Salah satu keunggulan utama dari arsitektur ini adalah kemampuannya untuk menangani data dalam skala besar. Amazon Kinesis Data Firehose memungkinkan pengumpulan data secara real-time dari Twitter API, sementara Amazon S3 menyediakan penyimpanan yang scalable dan aman untuk data mentah maupun hasil analisis. Dengan menggunakan layanan tanpa server seperti AWS Lambda, sistem ini dapat secara otomatis menyesuaikan kapasitas pemrosesan sesuai dengan volume data yang masuk, tanpa perlu khawatir tentang manajemen infrastruktur.

5.2. Real-Time Processing

Arsitektur ini dirancang untuk memproses data secara real-time, yang sangat penting dalam konteks analisis sentimen di media sosial. Data tweet yang dikumpulkan dari Twitter API langsung diproses dan dianalisis menggunakan Amazon Comprehend, yang mampu mengidentifikasi sentimen (positif, negatif, atau netral) serta entitas yang disebutkan dalam tweet. Hasil analisis ini kemudian dapat segera diakses oleh pengguna bisnis melalui Amazon QuickSight, memungkinkan mereka untuk mengambil keputusan berdasarkan informasi terkini.

5.3. Integrasi yang Mudah

Salah satu kelebihan dari menggunakan layanan AWS adalah kemudahan integrasi antara berbagai komponen. Amazon Kinesis terintegrasi dengan baik dengan Amazon S3, sementara AWS Lambda dapat dipicu oleh event S3 untuk memproses data secara otomatis. Amazon Comprehend dan Amazon Athena juga terintegrasi dengan baik, memungkinkan analisis data dan query SQL dilakukan dengan mudah. Terakhir, Amazon QuickSight menyediakan visualisasi data yang interaktif, sehingga pengguna bisnis dapat dengan mudah memahami hasil analisis.

5.4. Manfaat bagi Bisnis

Arsitektur ini memberikan manfaat yang signifikan bagi bisnis, terutama dalam hal pengambilan keputusan yang berbasis data. Dengan menganalisis sentimen dari tweet, perusahaan dapat memahami bagaimana pelanggan merespons produk atau layanan mereka, mengidentifikasi tren pasar, dan merespons keluhan pelanggan dengan cepat. Selain itu, sistem ini juga dapat digunakan untuk memantau kampanye pemasaran, mengukur efektivitas iklan, dan mengidentifikasi peluang baru di pasar.

5.5. Potensi Pengembangan

Meskipun arsitektur ini sudah sangat efektif, masih ada ruang untuk pengembangan lebih lanjut. Misalnya, sistem ini dapat diperluas untuk menganalisis data dari platform media sosial lainnya seperti Instagram atau Facebook. Selain itu, integrasi dengan layanan Machine Learning lainnya seperti Amazon SageMaker dapat meningkatkan akurasi dan kedalaman analisis sentimen. Penggunaan Amazon Translate juga dapat diperluas untuk mendukung lebih banyak bahasa, sehingga analisis sentimen dapat dilakukan secara global.

5.6. Kesimpulan Akhir

Secara keseluruhan, arsitektur Big Data untuk analisis sentimen Twitter menggunakan layanan AWS adalah solusi yang powerful, scalable, dan efisien. Dengan memanfaatkan layanan AWS, perusahaan dapat membangun sistem yang mampu mengolah data dalam skala besar, memprosesnya secara real-time, dan menghasilkan insights yang berguna untuk pengambilan keputusan bisnis. Arsitektur ini tidak hanya cocok untuk analisis sentimen Twitter, tetapi juga dapat diaplikasikan pada berbagai kasus penggunaan lainnya, menjadikannya investasi yang berharga bagi perusahaan di era *data-driven* seperti sekarang.

BAB 6: Sumber dan Referensi

- <https://docs.aws.amazon.com/pdfs/whitepapers/latest/big-data-analytics-options/big-data-analytics-options.pdf#example-3-sentiment-analysis-of-social-media>
- <https://aws.amazon.com/firehose/>
- <https://aws.amazon.com/s3/>
- <https://aws.amazon.com/lambda/>
- <https://aws.amazon.com/comprehend/>
- <https://aws.amazon.com/athena/>
- <https://aws.amazon.com/quicksight/>
- <https://aws.amazon.com/translate/>
- <https://aws.amazon.com/glue/>
- <https://aws.amazon.com/blogs/machine-learning/performing-sentiment-analysis-on-twitter-data-with-amazon-comprehend/>
- <https://aws.amazon.com/blogs/big-data/building-a-big-data-architecture-on-aws/>
- <https://aws.amazon.com/blogs/compute/real-time-data-processing-with-amazon-kinesis-and-aws-lambda/>
- <https://aws.amazon.com/blogs/big-data/visualizing-data-with-amazon-quicksight/>
- <https://developer.twitter.com/en/docs>
- <https://towardsdatascience.com/natural-language-processing-with-amazon-comprehend-7a8f2a0a9e5a>
- <https://www.datacouncil.ai/blog/big-data-architecture-best-practices>
- <https://medium.com/aws-activate-startup-blog/introduction-to-amazon-athena-9f5c0c8b5f5b>
- <https://www.serverless.com/aws-lambda>
- <https://www.amazon.com/Big-Data-Principles-Practices-Perspectives/dp/0134291077>
- <https://www.amazon.com/Data-Science-AWS-Implementing-End-to-End/dp/1492079391>
- <https://www.youtube.com/watch?v=7m6Zk6l0w5k>
- <https://www.youtube.com/watch?v=5x5X5X5X5X5>