# Proposal

**IF5270 Big Data System**

**Topik Analitik Berdasarkan Spotify API**



Oleh:

|  |  |
| --- | --- |
| Group 5  Brillian Fahmi Fauzi | 235210 |
| Iwan Kurniawan | 23521058 |
| Hamzah Misbachul | 235220 |

**Program Studi Magister Informatika Sekolah Teknik Elektro dan Informatika Institut Teknologi Bandung**

**2023**

1. **Latar Belakang**

Data Spotify API adalah data yang dapat diakses secara langsung dari Spotify, salah satu layanan streaming musik terbesar di dunia. Data Spotify API berisi informasi mengenai lagu, album, artis, playlist, dan sebagainya. Beberapa contoh data yang tersedia di Spotify API antara lain:

* Informasi mengenai lagu, seperti judul, artis, durasi, album, dan genre.
* Informasi mengenai artis, seperti nama, popularitas, dan genre.
* Informasi mengenai playlist, seperti judul, jumlah lagu, dan deskripsi.

# Persoalan

Berdasarkan latar belakang yang telah disebutkan, waktu yang dihabiskan untuk listing data dari spotify membutuhkan waktu. Maka dari itu, dalam Tugas Besar ini akan dibuat sebuah sistem Topik Analitik berdasarkan API pada Spotify.

# Tantangan

Dalam Tugas Besar ini, terdapat beberapa tantangan sebagai berikut:

* 1. Autentikasi: Untuk mengakses data Spotify, Kita harus memiliki akses token atau otentikasi. Token ini dapat berubah setiap saat, sehingga kita harus memperbarui token secara teratur agar tetap dapat mengakses data.
  2. Rate limiting: Spotify memiliki batasan tarif yang membatasi jumlah permintaan API yang dapat dilakukan dalam satu waktu. Jika kita melakukan permintaan terlalu sering atau terlalu banyak dalam waktu singkat, kita bisa dibatasi dan tidak dapat mengakses data selama beberapa waktu.
  3. Pengolahan Data yang Besar: Spotify memiliki miliaran lagu dan artis yang tersedia, dan memproses semua data tersebut dalam waktu yang singkat bisa sangat menantang. Oleh karena itu, memilih teknik pengolahan data yang tepat dan mengoptimalkan kinerja kode kita menjadi sangat penting.
  4. Perubahan dalam API: Spotify secara teratur memperbarui API mereka, sehingga kode yang telah ditulis sebelumnya mungkin tidak lagi berfungsi dengan benar. Kita harus memperbarui kode kita secara berkala untuk memastikan bahwa kita selalu dapat mengakses data terbaru.
  5. Perlindungan Hak Cipta: Dalam mengambil data dari Spotify, kita perlu memastikan bahwa kita tidak melanggar hak cipta mereka. kita tidak dapat mengambil data tanpa izin atau memperbanyak data yang dilindungi hak cipta tanpa izin dari pemilik hak cipta.

# Konsep Sistem

Kami membuat 3 layer pada sistem ini, yaitu layer staging, layer processing dan layer analytic.

Gambar Sistem

Konsep Big Data dapat diterapkan pada pengolahan data Spotify menggunakan Kafka. Pada umumnya, Big Data mengacu pada jumlah data yang sangat besar dan kompleks yang sulit untuk diolah menggunakan metode pengolahan data tradisional. Apache Kafka memungkinkan pengolahan data streaming secara real-time, yang memungkinkan pengolahan data yang cepat dan tepat waktu. Beberapa contoh penerapan Kafka dalam pengolahan data Spotify adalah sebagai berikut:

1. Streaming data dari Spotify ke Kafka: Data Spotify dapat diproses dalam waktu nyata menggunakan Kafka. Kafka memungkinkan streaming data dari Spotify ke broker Kafka. Dalam hal ini, Kafka dapat digunakan untuk mengambil informasi seperti informasi tentang lagu, album, dan artis, serta informasi lainnya dari data Spotify.
2. Memproses data yang besar: Data Spotify sangat besar dan kompleks, oleh karena itu Kafka dapat digunakan untuk memproses data yang besar tersebut. Kafka memungkinkan mempartisi data ke dalam beberapa partisi, yang dapat diproses secara terpisah dan dijalankan secara terdistribusi.
3. Integrasi dengan alat pengolahan Big Data lainnya: Kafka dapat diintegrasikan dengan alat pengolahan Big Data seperti Hadoop, Spark, dan Flink. Hal ini memungkinkan untuk melakukan analisis data terhadap data yang telah diambil dari Spotify menggunakan Kafka.
4. Penerapan fitur replikasi data: Kafka memiliki fitur replikasi data yang memungkinkan untuk membuat salinan data di beberapa server. Hal ini memungkinkan untuk meningkatkan keandalan data dan ketersediaan data.

Dalam keseluruhan, Kafka dapat digunakan sebagai platform streaming data yang handal, scalable, dan realtime dalam pengolahan data Spotify. Penggunaan Kafka dapat mempermudah pengolahan data yang besar dan kompleks dalam waktu nyata, yang sangat penting dalam pengolahan data big data.

**Konsep Big Data**

Gambaran umum tentang konsep tool big data dengan data Spotify:

1. Pengambilan Data: Data Spotify dapat diambil dari API Spotify atau dengan cara web scraping menggunakan library seperti BeautifulSoup. Data dapat diambil dalam format JSON atau CSV.
2. Penyimpanan Data: Data Spotify kemudian disimpan dalam sistem penyimpanan big data, seperti Hadoop Distributed File System (HDFS), Apache Cassandra, atau Amazon S3. Data yang disimpan dapat dianalisis secara cepat dan efisien menggunakan alat analisis data big data.
3. 1Proses Data: Data Spotify kemudian diproses menggunakan alat analisis big data, seperti Apache Spark atau Apache Flink. Alat-alat ini memungkinkan pengolahan data dalam skala besar dan real-time, yang memungkinkan pengolahan data Spotify secara efisien dan cepat.
4. Analisis Data: Setelah data diproses, data dapat dianalisis menggunakan alat analisis data big data, seperti Apache Hadoop atau Elasticsearch. Analisis data ini dapat digunakan untuk mengidentifikasi tren dan pola dalam perilaku pengguna Spotify dan membantu dalam membuat keputusan bisnis yang lebih baik.
5. Visualisasi Data: Hasil analisis data dapat di visualisasikan menggunakan alat visualisasi data big data seperti Apache Zeppelin atau Tableau. Hal ini memungkinkan untuk menyajikan hasil analisis data secara lebih intuitif dan mudah dimengerti.
6. Keputusan Bisnis: Hasil analisis data dapat digunakan untuk membuat keputusan bisnis yang lebih baik. Contohnya, hasil analisis data dapat digunakan untuk memahami preferensi pengguna Spotify dan membuat rekomendasi lagu atau playlist yang lebih baik.

Dalam keseluruhan, konsep tool big data dengan data Spotify melibatkan pengambilan, penyimpanan, pengolahan, analisis, visualisasi, dan pengambilan keputusan atas data Spotify. Penggunaan alat analisis data big data dapat membantu dalam pengolahan dan analisis data Spotify secara cepat, efisien, dan real-time, sehingga dapat membantu dalam mengambil keputusan bisnis yang lebih baik.

# Infrastruktur

# Dalam konteks Big Data, ada beberapa alat atau tool yang dapat digunakan untuk memproses data Spotify. Beberapa contoh tool big data yang umum digunakan dalam pengolahan data Spotify adalah sebagai berikut:

1. Apache Hadoop: Hadoop adalah kerangka kerja sumber terbuka yang digunakan untuk memproses data besar. Hadoop terdiri dari beberapa komponen, seperti HDFS (Hadoop Distributed File System) dan MapReduce, yang memungkinkan pengolahan data dalam skala besar secara terdistribusi. Hadoop dapat digunakan untuk memproses data Spotify dan melakukan analisis data seperti pengolahan log dan profil pengguna.
2. Apache Spark: Spark adalah kerangka kerja pengolahan data terdistribusi yang digunakan untuk memproses data besar dengan cepat. Spark mendukung pengolahan data dalam skala besar dan real-time, yang memungkinkan untuk memproses data Spotify secara cepat dan efisien.
3. Apache Flink: Flink adalah kerangka kerja pengolahan data terdistribusi yang mendukung pengolahan data real-time dan streaming. Flink dapat digunakan untuk memproses data Spotify secara real-time, sehingga data dapat dianalisis dan digunakan secara lebih cepat.
4. Elasticsearch: Elasticsearch adalah mesin pencari dan analitik data yang didesain untuk memproses data semi-struktur dan tidak terstruktur. Elasticsearch dapat digunakan untuk memproses data Spotify seperti pengolahan lagu, album, artis, dan playlist.
5. Apache Kafka: Seperti yang telah disebutkan sebelumnya, Kafka adalah platform pengolahan streaming data yang handal dan realtime. Kafka dapat digunakan untuk memproses data Spotify dalam waktu nyata, sehingga data dapat diakses dan digunakan dengan cepat.

Dalam keseluruhan, tool big data dapat digunakan untuk memproses dan menganalisis data Spotify dengan lebih cepat dan efisien. Alat-alat ini memungkinkan pengolahan data yang besar dan kompleks dalam skala besar dan real-time. Dalam hal ini, tool big data dapat membantu dalam mengambil keputusan bisnis yang lebih baik dan membuat strategi yang lebih efektif.

# Peran

|  |  |  |
| --- | --- | --- |
| **Nama** | **NIM** | **Peran** |
| Brillian Fahmi Fauzi | 235210 | Software Engineer |
| Iwan Kurniawan | 235210 | Software Engineer, Dokumentasi |
| Hamzah Misbachul | 235220 | BI Engineer |

1. **Metrics Evaluasi**

Metrik evaluasi yang digunakan dalam Tugas Besar ini adalah mengukur kecepatan pengambilan data Spotify dan pengolahan data tersebut. Semakin cepat data dapat diproses, semakin cepat informasi dapat dihasilkan.

# Referensi

# "Spotify Big Data Architecture and Recommendations" oleh Satyam Shivam Sundaram: Artikel ini memberikan gambaran tentang arsitektur dan rekomendasi untuk mengelola Big Data di Spotify.

# "Streaming Big Data with Spark Streaming & Spotify" oleh John Casey: Artikel ini membahas bagaimana menggunakan Spark Streaming untuk mengakses data Spotify dan mengelola Big Data.

# "Big Data Analytics with Spark and Hadoop for Spotify Music Recommendation" oleh Xuemei Yang dan Xiaodong Liu: Artikel ini membahas implementasi teknologi Big Data seperti Spark dan Hadoop untuk menganalisis data Spotify dan merekomendasikan musik.

# "Spotify: Big Data, Machine Learning, and Streaming" oleh Jack Vaughan: Artikel ini membahas bagaimana Spotify menggunakan Big Data dan Machine Learning untuk menyediakan layanan streaming musik yang personalisasi dan efisien.

# "Big Data Analytics with Hadoop and Hive for Spotify Music Recommendation" oleh Sandeep K. Gupta dan Puneet Kumar: Artikel ini membahas bagaimana menggunakan Hadoop dan Hive untuk menganalisis data Spotify dan memberikan rekomendasi musik berdasarkan preferensi pengguna.

# "Using Big Data to Create the Perfect Playlist" oleh Daphne Leprince-Ringuet: Artikel ini membahas bagaimana Spotify menggunakan Big Data untuk membuat playlist yang disesuaikan dengan preferensi pengguna.

# "Building Big Data Pipeline with Apache Kafka and Spotify" oleh Bhavuk Chawla: Artikel ini membahas penggunaan Apache Kafka untuk membangun pipeline Big Data di Spotify.