

INTEGRASI DAN INTEROPERABILITAS DALAM BIG DATA

MATA KULIAH INFRASTRUKTUR DAN TEKNOLOGI BIG DATA

OLEH: GALIH HERMAWAN, S.KOM., M.T.

PROGRAM STUDI TEKNIK INFORMATIKA. FTIK.

UNIVERSITAS KOMPUTER INDONESIA



INTEGRASI DATA DALAM BIG DATA

- Apa itu Integrasi Data?
 - Menggabungkan data dari berbagai sumber.
 - Meningkatkan nilai data.
 - Mendukung pengambilan keputusan.
- Dalam ekosistem Big Data, data berasal dari berbagai sumber dengan format yang berbeda-beda, menimbulkan tantangan integrasi.
- Dalam menangani *Volume*, *Velocity*, *Variety*, dan memastikan *Veracity*, integrasi data memainkan peran kunci untuk mengekstrak nilai (*Value*) dari Big Data.

LINGKUNGAN BIG DATA YANG HETEROGEN

- Beragamnya Sumber Data: Data dari basis data, *streaming*, fail teks, dan lainnya.
- Format Data yang Berbeda: JSON, CSV, XML, dan sebagainya.
- Tantangan Integrasi: Kesulitan menggabungkan data heterogen.

TANTANGAN INTEGRASI DATA

- Data dari berbagai sumber tersimpan dalam platform yang berbeda, sehingga sulit untuk disinkronisasikan dan diakses secara mudah.
- Data yang berkualitas rendah, tidak lengkap, tidak konsisten, atau duplikat, sehingga mempengaruhi hasil analisis dan pengambilan keputusan.
- Data yang terus bertambah dan berubah, sehingga memerlukan teknik pemrosesan yang cepat dan fleksibel.
- Data yang sensitif atau rahasia, sehingga memerlukan teknik keamanan dan privasi yang kuat.

KARAKTERISTIK BIG DATA 5V DAN INTEGRASI DATA

- **Volume:**
 - Karakteristik: Jumlah data yang sangat besar, terus meningkat, dan beragam.
 - Tantangan Integrasi: Menggabungkan data dari berbagai sumber dengan volume besar.
- **Velocity:**
 - Karakteristik: Kecepatan data yang dihasilkan dan bergerak sangat cepat.
 - Tantangan Integrasi: Menggabungkan data *streaming real-time* dengan data *batch* tradisional.
- **Variety:**
 - Karakteristik: Berbagai jenis data, seperti terstruktur, semi-terstruktur, dan tidak terstruktur.
 - Tantangan Integrasi: Menggabungkan data dengan format dan struktur yang berbeda.
- **Veracity:**
 - Karakteristik: Keakuratan dan keandalan data.
 - Tantangan Integrasi: Memastikan kualitas data dari berbagai sumber.
- **Value:**
 - Karakteristik: Nilai yang dapat diperoleh dari data.
 - Tantangan Integrasi: Mengubah data menjadi informasi yang bermanfaat.

MANFAAT INTEGRASI DATA

- **Meningkatkan kualitas data:** Integrasi data dapat membantu mengidentifikasi dan memperbaiki data yang tidak akurat atau tidak lengkap.
- **Meningkatkan efisiensi:** Integrasi data dapat mengotomatiskan proses pengumpulan dan transformasi data.
- **Meningkatkan skalabilitas:** Integrasi data dapat membantu organisasi untuk menangani volume data yang terus meningkat.
- **Meningkatkan visibilitas:** Integrasi data dapat memberikan pandangan yang lebih komprehensif dari data organisasi.
- **Meningkatkan nilai data:** Integrasi data dapat membantu organisasi untuk mendapatkan wawasan yang lebih berharga dari data.

TEKNOLOGI INTEGRASI DATA

- **Apache Nifi:**
 - Platform integrasi data yang *scalable*.
 - Mendukung berbagai sumber dan tujuan data.
 - Menyediakan berbagai komponen untuk transformasi data.
- **Apache Kafka:**
 - Platform *streaming* data yang *scalable*.
 - Mendukung *real-time* data *processing*.
 - Memudahkan integrasi data *streaming*.
- **Informatica PowerCenter:**
 - Alat integrasi data yang populer dan kuat.
 - Mendukung berbagai sumber dan tujuan data.
 - Menyediakan berbagai fitur untuk transformasi data, pembersihan data, dan manajemen kualitas data.
- **IBM InfoSphere DataStage:**
 - Alat integrasi data yang komprehensif dan skalabel.
 - Mendukung berbagai platform dan teknologi.
 - Menyediakan berbagai fitur untuk integrasi data *batch* dan *real-time*.

TEKNOLOGI INTEGRASI DATA (LANJUTAN)

- **Microsoft SQL Server Integration Services (SSIS):**
 - Alat integrasi data yang terintegrasi dengan Microsoft SQL Server.
 - Mudah digunakan dan dipelajari.
 - Mendukung berbagai sumber dan tujuan data.
- **Data Integrator (ODI):**
 - Alat integrasi data yang lengkap dan skalabel.
 - Mendukung berbagai platform dan teknologi.
 - Menyediakan berbagai fitur untuk integrasi data *batch* dan *real-time*.
- **SAP Data Services**
 - Alat integrasi data yang dirancang untuk lingkungan SAP.
 - Mendukung berbagai sumber dan tujuan data SAP.
 - Menyediakan berbagai fitur untuk transformasi data, pembersihan data, dan manajemen kualitas data.

TEKNIK INTEGRASI DATA

- **Extract, Transform, Load (ETL):**
 - Proses tradisional untuk integrasi data.
 - Menggabungkan data dari berbagai sumber.
 - Membersihkan dan mentransformasi data.
 - Memuat data ke dalam target data warehouse.
- **Real-time Data Integration:**
 - Integrasi data *streaming*.
 - Mendukung pengambilan keputusan *real-time*.
 - Membutuhkan teknologi seperti Apache Kafka.

INTEROPERABILITAS DATA

- Kemampuan berbagai sistem untuk berbagi data, saling berkomunikasi dan bekerja sama.
- Memungkinkan analisis data yang lebih komprehensif.
- Meningkatkan nilai data.
- Standar interoperabilitas:
 - XML.
 - JSON.
 - RDF.

TREN INTEGRASI DATA

- **Integrasi data *real-time***: Menjadi semakin penting untuk pengambilan keputusan yang cepat.
- ***Cloud computing***: Menyediakan *platform* yang *scalable* dan *cost-effective* untuk integrasi data.
- ***Artificial intelligence (AI)***: Digunakan untuk meningkatkan kualitas data dan mengotomatisasi proses integrasi data.

CONTOH PENERAPAN INTEGRASI DATA

- **Analisis ritel:** Menganalisis data pembelian untuk memahami tren dan meningkatkan rekomendasi produk.
- **Analisis layanan kesehatan:** Menganalisis data pasien untuk meningkatkan kualitas perawatan dan mengurangi biaya.
- **Analisis keuangan:** Menganalisis data pasar untuk mendeteksi peluang dan risiko.

RINGKASAN

- Big Data adalah teknologi dan teknik untuk memproses dan menganalisis data yang memiliki jumlah yang sangat besar, baik yang terstruktur, semi terstruktur dan tidak terstruktur.
- Integrasi data adalah proses untuk mengumpulkan semua data yang telah dibuat hingga menjadi Big Data, yang memiliki tantangan seperti data yang tersimpan dalam platform yang berbeda, data yang berkualitas rendah, data yang terus bertambah dan berubah, dan data yang sensitif atau rahasia.
- Teknologi integrasi data adalah alat atau platform yang digunakan untuk mengumpulkan, menyimpan, memproses, dan menganalisis data dari berbagai sumber, seperti Apache Nifi, Apache Kafka, dll.
- Teknik integrasi dan interoperabilitas adalah metode atau prosedur yang digunakan untuk menggabungkan dan menyelaraskan data dari berbagai sumber, sehingga dapat digunakan secara bersama-sama untuk tujuan analisis, seperti teknik ETL, data *mapping*, dan data *conversion*.

STUDI KASUS BIG DATA DAN INTEGRASI DATA

- **Netflix:** adalah sebuah perusahaan penyedia layanan *streaming* film dan serial TV yang menggunakan Big Data dan integrasi data untuk meningkatkan pengalaman pengguna. Netflix mengumpulkan data dari berbagai sumber, seperti data perilaku, data preferensi, data *rating*, data demografi, dan data lokasi pengguna. Data tersebut kemudian disimpan dalam data *warehouse* dan data *lake*, yang dapat diakses oleh berbagai platform dan aplikasi. Netflix kemudian memproses dan menganalisis data tersebut dengan menggunakan algoritma *machine learning* untuk memberikan rekomendasi film dan serial TV yang sesuai dengan selera pengguna, serta untuk mengoptimalkan kualitas *streaming*.
- **Starbucks:** adalah sebuah perusahaan kopi terkenal yang menggunakan Big Data dan integrasi data untuk meningkatkan loyalitas pelanggan. Starbucks mengumpulkan data dari berbagai sumber, seperti data transaksi, data aplikasi, data media sosial, data cuaca, dan data lalu lintas. Data tersebut kemudian disimpan dalam data *lake*, yang dapat diakses oleh berbagai platform dan aplikasi. Starbucks kemudian memproses dan menganalisis data tersebut dengan menggunakan teknik analitik untuk memberikan penawaran dan promosi yang sesuai dengan kebutuhan dan minat pelanggan, serta untuk meningkatkan efisiensi operasional.
- **Spotify:** adalah sebuah perusahaan penyedia layanan *streaming* musik dan *podcast* yang menggunakan Big Data dan integrasi data untuk meningkatkan kualitas layanan. Spotify mengumpulkan data dari berbagai sumber, seperti data lagu, data artis, data *playlist*, data pengguna, dan data sensor. Data tersebut kemudian disimpan dalam data *lake*, yang dapat diakses oleh berbagai platform dan aplikasi. Spotify kemudian memproses dan menganalisis data tersebut dengan menggunakan teknik *machine learning* untuk memberikan rekomendasi lagu dan *podcast* yang sesuai dengan selera pengguna, serta untuk menghasilkan fitur personalisasi, seperti Discover Weekly dan Wrapped.

STUDI KASUS - INTEGRASI DATA UNTUK MENINGKATKAN PENGALAMAN PELANGGAN

- **Contoh:**

- Sebuah perusahaan ritel besar ingin meningkatkan pengalaman pelanggannya. Mereka memiliki data tentang pembelian pelanggan, interaksi layanan pelanggan, dan data demografis. Namun, data ini tersebar di berbagai sistem dan tidak terintegrasi.

- **Solusi:**

Perusahaan menerapkan platform integrasi data untuk menggabungkan data dari berbagai sumber. Ini memungkinkan mereka untuk:

- Menganalisis perilaku pelanggan: Mereka dapat mengidentifikasi pola pembelian dan preferensi pelanggan.
- Mempersonalisasi penawaran: Mereka dapat memberikan rekomendasi produk yang lebih relevan kepada pelanggan.
- Meningkatkan layanan pelanggan: Mereka dapat mengidentifikasi dan menyelesaikan masalah pelanggan dengan lebih cepat.



STUDI KASUS (LANJUTAN)

- **Hasil:**
 - Integrasi data membantu perusahaan untuk meningkatkan kepuasan pelanggan dan meningkatkan penjualan.
- **Teknologi Integrasi Data yang Digunakan:**
 - Apache Nifi: Digunakan untuk mengumpulkan dan memindahkan data dari berbagai sumber.
 - Apache Kafka: Digunakan untuk memproses data streaming real-time dari interaksi layanan pelanggan.
 - Apache Hive: Digunakan untuk menyimpan dan menganalisis data terstruktur.

STUDI KASUS (LANJUTAN)

- **Wawasan:**

Integrasi data dapat membantu organisasi untuk:

- Memperoleh wawasan yang lebih komprehensif dari data: Dengan menggabungkan data dari berbagai sumber, organisasi dapat melihat gambaran yang lebih lengkap tentang pelanggan, operasi, dan pasar mereka.
- Meningkatkan pengambilan keputusan: Wawasan yang diperoleh dari data terintegrasi dapat membantu organisasi untuk membuat keputusan yang lebih baik dan lebih cepat.
- Meningkatkan peluang bisnis: Integrasi data dapat membantu organisasi untuk mengidentifikasi peluang baru untuk meningkatkan produk, layanan, dan operasi mereka.

- **Kesimpulan:**

- Integrasi data adalah alat yang ampuh yang dapat membantu organisasi untuk mendapatkan nilai dari Big Data. Dengan mengintegrasikan data dari berbagai sumber, organisasi dapat memperoleh wawasan yang lebih komprehensif, meningkatkan pengambilan keputusan, dan meningkatkan peluang bisnis.

- **Catatan:**

- Studi kasus ini hanya contoh dan dapat dimodifikasi sesuai kebutuhan.
- Anda dapat menambahkan lebih banyak detail dan contoh untuk memperkaya studi kasus.

DISKUSI

TERIMA KASIH