

Nama: Qorina M. H. Mumtaza

Kelas: DE-1

1. Apa peran utama seorang Data Engineer dalam ekosistem data? Bagaimana peran ini berbeda dari Data Scientist dan Data Analyst?

Jawab:

Data Engineer memainkan peran krusial di ekosistem data dengan berfokus pada desain, konstruksi, instalasi, dan pemeliharaan sistem yang memungkinkan pengumpulan, penyimpanan, serta pengambilan data. Tanggung jawab utamanya termasuk:

- **Membuat Data Pipeline:** Data Engineer membangun dan memelihara data pipeline yang mengumpulkan, membersihkan, dan mentransformasi data dari berbagai sumber, sehingga menjadikannya siap untuk dianalisis. Pipeline ini seringkali melibatkan ekstraksi data dari berbagai database, API, logs, dan sumber lainnya.
- **Data Warehousing:** Data Engineer mendesain dan mengelola data warehouse yang merupakan repositori lebih besar untuk menyimpan data terstruktur agar lebih mudah diakses dan dianalisis oleh anggota tim lain, seperti Data Scientist dan Data Analyst.
- **Database Management:** Data Engineer bekerja dengan system database yang bervariasi untuk memastikan data disimpan dengan aman, efisien dan dapat diakses sewaktu-waktu ketika dibutuhkan.
- **Data Quality and Governance:** Data Engineer bertanggung jawab untuk memastikan kualitas data dan menetapkan kebijakan tata kelola data untuk menjaga integritas dan keamanan data.
- **Scalability and Performance:** Data Engineer bekerja untuk memastikan system data dapat mengatasi peningkatan volume data dan bekerja secara efisien, bahkan ketika kebutuhan data meningkat.
- **Infrastructure Management:** Data Engineer seringkali mengelola infrastruktur yang mendasar untuk mendukung penyimpanan dan pemrosesan data.
- **Tool and Technology Selections:** Data Engineer mengevaluasi dan memilih tools dan teknologi yang cocok untuk penyimpanan, pemrosesan, dan analisis data berdasarkan kebutuhan spesifik.

Dari beberapa tanggung jawab utama yang telah disebutkan sebelumnya, tentu Data Engineer memiliki peran yang berbeda dengan Data Scientist dan Data Analyst. Data Engineer cenderung fokus pada infrastruktur dan tools yang diperlukan untuk pengumpulan, penyimpanan, dan pemrosesan data. Sedangkan Data Scientist utamanya bertanggung jawab atas ekstraksi nilai dari data. Mereka cenderung melakukan analisis statistik dan analisis machine learning, membangun model prediksi, dan membuat rekomendasi berdasarkan data untuk melakukan pengambilan keputusan. Data Scientist seringkali mengandalkan data yang disiapkan oleh Data Engineer.

Begitu halnya dengan Data Analyst sebagai konsumen data, mereka melakukan query database, menghasilkan laporan, membuat visualisasi data, dan melakukan analisis deskriptif dasar. Data Analyst seringkali bekerja dengan data terstruktur dan menggunakan alat yang dibuat oleh Data Engineer.

2. Berikan beberapa contoh peranan dari seorang Data Engineer yang mungkin bersinggungan atau bahkan sama dengan peran Data Scientist dan Data Analyst!

Jawab:

Meskipun Data Engineer, Data Scientist, dan Data Analyst memiliki tanggung jawab utama yang berbeda, terdapat area yang tumpang tindih dan tugas yang harus dibagi karena sifatnya kolaboratif, seperti:

- **Data Cleaning dan Transformasi:** Data Engineer biasanya melakukan pembersihan dan transformasi data awal untuk menyiapkan data mentah untuk kemudian dianalisis. Data Scientist juga dapat terlibat dalam pembersihan dan transformasi data ketika mereka perlu menyesuaikan data dengan persyaratan pemodelan atau analisis tertentu. Sedangkan Data Analyst seringkali membersihkan dan mengubah data untuk menghasilkan insight dan membuat visualisasi untuk laporan.
- **Data Exploration and Discovery:** Data Engineer dapat mengeksplorasi data untuk mengidentifikasi masalah atau anomali kualitas data sebagai bagian dari pengembangan saluran data. Data Scientist mengeksplorasi data selama tahap awal proyek untuk memahami karakteristiknya dan mengidentifikasi pola atau fitur potensial untuk membangun pemodelan. Sedangkan Data Analyst mengeksplorasi data untuk mengetahui tren, pola, dan anomaly yang mana akan dibuat laporan dan visualisasi.
- **Database Querying and Reporting:** Data Engineer menulis dan mengoptimalkan query basis data yang mengekstrak data untuk berbagai tujuan analisis dan pelaporan. Data Scientist seringkali membuat query database untuk mengakses kumpulan data spesifik untuk dianalisis. Sedangkan Data Analyst sangat bergantung pada query dan pelaporan basis data untuk menghasilkan wawasan dan membuat dashboard.

3. Jelaskan langkah-langkah proses ETL dan ELT yang berperan dalam pekerjaan seorang Data Engineer!

- **ETL (Extract > Transform > Load)**
 - Extract** : Mengambil data mentah dari kumpulan data tidak terstruktur dan memindahkannya ke stase penyimpanan data yang temporer.
 - Clean** : Membersihkan data yang sudah diekstraksi dari kumpulan data tidak terstruktur untuk memastikan kualitas data sebelum ditransformasi.
 - Transform** : Strukturisasi dan mengkonversi data untuk mencocokkan sumber target yang benar.
 - Load** : Memuat (*me-load*) data terstruktur ke gudang data (*data warehouse*) atau data lake sehingga dapat dianalisa dan digunakan dengan benar.

Analyze : Analisis big data diproses dalam gudang data, memungkinkan bisnis untuk mendapatkan *insight* baru dari data yang dikonfigurasi dengan benar.

- ELT (Extract > Load > Transform)

Extract : Langkah ini memiliki cara kerja yang mirip dengan proses extract di ETL, sama-sama mengambil data mentah dari infrastruktur virtual dan memindahkannya ke stase penyimpanan data yang temporer.

Load : Pada langkah ini baru terdapat perbedaan dari proses ETL. ETL mengirim data mentah dan memuatnya ke server sementara untuk melalui proses transformasi, sedangkan ELT mengirimkan data langsung ke target lokasi penyimpanan. Proses ini mempersingkat siklus antara ekstraksi dan pengiriman data.

Transform : Database atau gudang data (*data warehouse*) menyortir dan menormalisasi data, menyimpan dan membuatnya mudah diakses untuk kegiatan pelaporan.