

PART 1 – Fundamental DE

TASK 1

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

Jawabannya : Peran seorang Data Engineer dalam ekosistem data sangat penting. Mereka bertanggung jawab untuk merancang, membangun, dan menjaga infrastruktur teknis yang memungkinkan pengumpulan, penyimpanan, dan analisis data yang efisien dan efektif.

Berikut beberapa peran utama seorang Data Engineer:

1. **Pengumpulan Data:** Data Engineer bertanggung jawab untuk merancang dan mengimplementasikan proses pengumpulan data dari berbagai sumber seperti basis data, sistem streaming, file teks, atau API.
2. **Pembersihan dan Pemrosesan Data:** Setelah data dikumpulkan, Data Engineer membersihkan dan memprosesnya untuk memastikan konsistensi, keakuratan, dan kelayakan untuk analisis lebih lanjut. Ini dapat melibatkan normalisasi data, menghilangkan duplikat, dan menangani data yang hilang atau rusak.
3. **Penyimpanan Data:** Data Engineer merancang dan mengelola infrastruktur penyimpanan data yang sesuai untuk memastikan data tersedia, aman, dan dapat diakses dengan cepat. Ini bisa mencakup basis data relasional, basis data NoSQL, sistem penyimpanan berbasis cloud, atau data warehouse.
4. **Pemrosesan Data:** Mereka juga bertanggung jawab untuk merancang dan mengelola proses pemrosesan data seperti pengolahan batch dan streaming. Hal ini memungkinkan data untuk diolah secara real-time atau dalam batch tergantung kebutuhan bisnis.
5. **Optimisasi Kinerja:** Data Engineer terus-menerus mengoptimalkan kinerja sistem data dengan meningkatkan efisiensi penyimpanan, mempercepat proses pengolahan data, dan mengoptimalkan kueri untuk memenuhi kebutuhan bisnis.

Perbedaan peran Data Engineer, Data Scientist, dan Data Analyst terletak pada fokus dan tanggung jawab masing-masing:

- **Data Engineer** lebih berfokus pada pembangunan dan pemeliharaan infrastruktur data. Data Engineer berurusan dengan pengumpulan, penyimpanan, dan pemrosesan data secara teknis.
- **Data Scientist** berfokus pada analisis dan interpretasi data untuk menghasilkan wawasan bisnis. Mereka menggunakan teknik statistik, machine learning, dan analisis data lanjutan untuk mengidentifikasi tren, membuat prediksi, dan mengambil keputusan yang didasarkan pada data.
- **Data Analyst** juga berfokus pada analisis data, tetapi dengan lebih sedikit fokus pada pemeliharaan infrastruktur data. Mereka cenderung menggunakan alat analisis dan kueri untuk menjawab pertanyaan bisnis spesifik dan menghasilkan laporan yang dapat dimengerti secara luas oleh pemangku kepentingan bisnis.

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

Jawaban : ada beberapa area di mana peran seorang Data Engineer mungkin bersinggungan atau bahkan sama dengan peran Data Scientist dan Data Analyst. Berikut adalah beberapa contohnya:

- **Pembersihan dan Pemrosesan Data (Cleaning):** Data Engineer bertanggung jawab untuk membersihkan dan memproses data mentah agar cocok untuk analisis lebih lanjut. Namun, Data Scientist juga harus memiliki keterampilan dalam pembersihan dan pemrosesan data karena kualitas data yang baik sangat penting untuk analisis yang akurat. Data Analyst juga mungkin terlibat dalam langkah-langkah awal ini untuk memastikan data yang digunakan untuk laporan dan analisis bersih dan berkualitas.
- **Pengolahan Data Real-Time:** Data Engineer biasanya bertanggung jawab untuk merancang dan mengelola infrastruktur untuk pemrosesan data real-time. Namun, Data Scientist juga dapat terlibat dalam pengolahan data real-time saat mereka melakukan analisis secara langsung pada data yang masuk untuk mendapatkan wawasan real-time. Data Analyst juga mungkin memerlukan akses ke data real-time untuk membuat laporan dan analisis yang diperbarui secara teratur.
- **Optimisasi Kinerja:** Data Engineer bertanggung jawab untuk mengoptimalkan kinerja infrastruktur data. Namun, Data Scientist dan Data Analyst juga harus memperhatikan kinerja saat mereka melakukan analisis data. Mereka dapat memberikan masukan kepada Data Engineer tentang kebutuhan kinerja tertentu untuk memastikan infrastruktur dapat memenuhi kebutuhan analisis mereka.
- **Pengembangan dan Penerapan Algoritma:** Data Scientist biasanya memiliki peran utama dalam mengembangkan dan menerapkan model machine learning dan algoritma analisis data lainnya. Namun, Data Engineer juga dapat terlibat dalam membangun infrastruktur yang diperlukan untuk mengelola dan menyimpan model yang dikembangkan oleh Data Scientist. Data Analyst mungkin menggunakan output dari algoritma ini dalam analisis mereka.
- **Kolaborasi Tim:** Meskipun tanggung jawab utama masing-masing peran berbeda, kolaborasi antara Data Engineer, Data Scientist, dan Data Analyst seringkali penting. Mereka perlu bekerja sama untuk memahami kebutuhan bisnis, memastikan data yang tersedia dan bersih, dan menghasilkan wawasan yang berarti dari data. Ini dapat melibatkan diskusi tentang infrastruktur data, metode analisis, dan interpretasi hasil.

3. Jelaskan langkah-langkah proses ETL dan ELT yang berperan dalam pekerjaan seorang data engineer !

Jawabannya : Proses ETL (Extract, Transform, Load) dan ELT (Extract, Load, Transform) adalah dua pendekatan yang umum digunakan dalam pengolahan data, terutama oleh seorang Data Engineer. Berikut langkah-langkah proses ETL dan ELT beserta peran mereka dalam pekerjaan seorang Data Engineer:

Extract (Ekstraksi):

ETL: Proses ekstraksi dalam ETL melibatkan pengambilan data dari berbagai sumber, seperti database, file teks, sumber streaming, API, dan lain-lain. Data Engineer bertanggung jawab untuk mengekstrak data secara efisien dari sumber-sumber ini.

ELT: Dalam ELT, ekstraksi juga merupakan langkah pertama, di mana data diekstraksi dari sumber-sumber yang sama seperti dalam ETL. Perbedaannya adalah bahwa data tersebut kemudian dimuat langsung ke penyimpanan data tanpa transformasi awal.

Transform (Transformasi):

ETL: Setelah data diekstraksi, langkah selanjutnya adalah mentransformasikannya menjadi format yang cocok untuk analisis. Ini bisa melibatkan pembersihan data, penggabungan data dari beberapa sumber, normalisasi, penghapusan data duplikat, atau pembentukan agregat. Data Engineer bertanggung jawab untuk merancang dan menerapkan proses transformasi ini.

ELT: Dalam ELT, langkah transformasi dilakukan setelah data dimuat ke penyimpanan data. Ini memungkinkan transformasi yang lebih fleksibel dan skala, karena data tetap dalam lingkungan yang dapat diakses.

Load (Pemuatan):

ETL: Setelah data diubah, langkah terakhir adalah memuatnya ke sistem penyimpanan data yang cocok, seperti data warehouse atau data lake. Data Engineer bertanggung jawab untuk memastikan bahwa proses pemuatan data berjalan lancar dan data tersedia untuk digunakan oleh pengguna akhir.

ELT: Dalam ELT, data dimuat terlebih dahulu ke penyimpanan data tanpa transformasi awal. Setelah itu, transformasi dapat diterapkan secara fleksibel sesuai kebutuhan analisis.

Peran seorang Data Engineer dalam proses ETL dan ELT meliputi merancang dan mengelola infrastruktur teknis untuk mendukung kedua pendekatan tersebut, mengembangkan dan mengimplementasikan skrip atau alur kerja untuk mengekstrak, mentransformasi, dan memuat data, memantau kinerja proses ETL atau ELT, serta melakukan pemeliharaan dan pembaruan terhadap sistem untuk memastikan kelancaran operasi. Selain itu, Data Engineer juga harus berkolaborasi dengan Data Scientist dan Data Analyst untuk memahami kebutuhan mereka dalam pengolahan data dan menyediakan solusi yang sesuai.