

Nama : Kharisma Novi Chandramukti

Kelas/Batch : Data Engineering 4

Part 1 – Fundamental DE

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

JAWAB :

- Peran seorang Data Engineer dalam ekosistem data adalah bertanggung jawab untuk merancang, membangun, mengelola dan mengoptimalkan infrastruktur data dan arsitektur untuk pengolahan data dalam skala yang besar
- Perbedaan Data Engineer dengan Data Scientist dan Data Analyst :
 - o Data Scientist berfokus pada analisis data untuk mendapatkan pemahaman yang mendalam dari data, dengan menggunakan berbagai algoritma dan teknik statistik serta machine learning untuk membuat prediksi, mengidentifikasi tren dan membuat rekomendasi.
 - o Data analyst berfokus untuk menganalisis data untuk menemukan pola, tren dan wawasan yang dapat mendukung pengambilan keputusan dalam bisnis, dengan menggunakan teknik analisis statistik dan visualisasi data untuk menjawab pertanyaan bisnis yang spesifik.

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

JAWAB :

Contoh peran dari seorang Data Engineer yang mungkin bersinggungan atau bahkan sama dengan peran Data Scientist dan Data Analyst :

- a. Pengelola data (Data Steward) : bertanggung jawab untuk memastikan kualitas data, metadata, dan dokumentasi data
- b. Ahli pengolahan big data : fokus pada pemrosesan data skala besar menggunakan teknologi seperti cloud
- c. Spesialis basis data : bertanggung jawab untuk mengelola basis data, membangun skema basis data, dan memastikan kinerja yang optimal

d. Pengembang algoritma : membuat megoptimalan algoritma untuk analisis data dan machine learning

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

JAWAB :

- ETL (Extract, Transform, Load), langkah – langkah dalam proses ETL sebagai berikut :
 1. Extract : data engineer mengekstrak data dari berbagai sumber, seperti basis data transaksional, file CSV, API, dll. Ekstraksi ini dapat dilakukan secara periodic atau secara real-time tergantung pada kebutuhan bisnis
 2. Transform : mencakup pembersihan data (data cleansing), penggabungan data dari berbagai sumber, penghapusan data yang tidak relevan, dan transformasi struktur data untuk memenuhi kebutuhan analisis yang akan dilakukan
 3. Load : memuat data ke dalam system data Gudang atau data mart yang sesuai
- ELT (Extract, Load, Transform), langkah – langkah dalam proses ELT sebagai berikut :
 1. Extract : diekstrak dari sumber data seperti pada ETL
 2. Load : dimuat ke dalam system penyimpanan data yang mentah (raw data storage) tanpa transformasi awal yang signifikan
 3. Transform : transformasi data dilakukan setelah data dimuat ke dalam sistem penyimpanan data. Data engineer menggunakan teknik seperti SQL atau alat pemrosesan big data untuk mentransformasi data sebelum menganalisisnya