Nurul Taufik Kurrahman

**Fundamental Data Engineer**

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

Data Engineer itu seperti chef di dapur yang menyiapkan bahan-bahan untuk dimasak.

Peran utama mereka adalah:

* Ekstraksi dan Transformasi Data: Mereka mengumpulkan data dari berbagai sumber, seperti database, aplikasi, atau sensor, terus mereka cleaning, transformasi, dan mempersiapkan data tersebut untuk bisa digunakan lebih lanjut.
* Mengelola Data Pipeline: Data Engineer punya tugas membuat sistem untuk memindahkan data dari satu tempat ke tempat lain dengan aman dan efisien. Jadi, data bisa bergerak mulus di dalam organisasi.
* Mengatur Data Warehouse: Mereka mengurus Data Warehouse dan sistem penyimpanan data. Memastikan data tersimpan dengan aman dan bisa diakses dengan cepat.

Sekarang, perbedaannya sama Data Scientist dan Data Analyst. Data Scientist itu seperti ilmuwan eksperimen. Mereka eksplorasi data, cari pola atau tren, dan membuat model prediksi. Mereka butuh data yang bersih dan terstruktur, yang biasanya sudah disiapin oleh Data Engineer.

Sementara itu, Data Analyst seperti detektif. Mereka menganalisis data yang udah ada untuk ngasih wawasan bisnis. Mereka lebih fokus ke laporan, visualisasi, dan pemahaman data yang udah ada. Biasanya, Data Analyst juga butuh Data Engineer untuk menyiapkan data buat mereka.

Jadi, Data Engineer itu seperti juru masak data, yang siapin bahan baku. Data Scientist itu eksperimen sama bahan baku itu buat hasilin resep makanan yang enak. Dan Data Analyst itu tukang nikmatin makanan yang udah jadi dan ceritain rasanya. Semua punya peran penting dalam ekosistem data.

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

* Data Cleaning: Data Engineer bisa membersihkan data mentah untuk menghilangkan nilai-nilai yang hilang, duplikat, atau tidak valid. Data Scientist dan Data Analyst juga perlu melakukan pembersihan data serupa sebelum analisis.
* Data Transformation: Data Engineer melakukan transformasi data untuk mengubah format, struktur, atau unit pengukuran data. Data Scientist atau Data Analyst mungkin memerlukan data yang sudah diubah ini untuk analisis lebih lanjut.
* Data Visualization: Data Engineer membuat dashboard analisa untuk data yang ada sebagai gambaran sementara sebelum dianalisa lebih lanjut oleh Data Analyst.

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

**ETL (Extract, Transform, Load):**

* Ekstraksi (Extract):

Tahap pertama adalah mengambil data dari berbagai sumber, seperti database, file teks, API, atau sistem eksternal. Data Engineer akan mengekstrak data mentah dari sumber ini dan memasukkannya ke dalam sistem ETL.

* Transformasi (Transform):

Di tahap transformasi, Data Engineer membersihkan dan mentransformasi data agar sesuai dengan kebutuhan. Ini mencakup menghapus data yang tidak valid, mengisi nilai yang hilang, menggabungkan data dari berbagai sumber, atau mengubah format data proses ini dilakukan di di penyimpanan sementara/staging area.

* Penyimpanan (Load):

Tahap terakhir adalah memuat data yang sudah diolah ke dalam sistem penyimpanan data, seperti data lake atau data warehouse yang dioptimalkan untuk analisis.

**ELT (Extract, Load, Transform):**

* Ekstraksi (Extract):

Seperti pada ETL, data masih diekstrak dari berbagai sumber, tetapi pada ELT, data diekstrak dan disimpan langsung ke dalam data lake terlebih dahulu dalam bentuk mentah, tanpa transformasi awal.

* Penyimpanan (Load):

Setelah data raw dimuat ke dalam data lake, tahap berikutnya adalah mengatur data ini ke dalam data waarehouse atau penyimpanan yang sesuai untuk analisis.

* Transformasi (Transform):

Transformasi data terjadi setelah data mentah dimuat. Dalam konteks ELT, transformasi lebih terfokus pada analisis yang dilakukan oleh Data Scientist atau Data Analyst dan proses ini dilakukan di data warehouse.