Part 1 - Fundamental DE

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

Deskripsi pekerjaan utama Data Engineer adalah menyediakan dan menyiapkan data agar bisa dikonsumsi oleh divisi lain dalam suatu project atau disebut menyiapkan data pipeline yang terdiri dari data collecting, processing, dan warehousing. Perbedaan peran ini dari data analyst dan scientist adalah, output dari pekerjaan data engineer nantinya akan dikonsumsi oleh data analyst untuk mencari insight-insight tertentu. Sedangkan data scientist lebih ke arah pembelajaran mesin untuk kebutuhan forecasting atau prediksi, deteksi, dll.

1. Berikan beberapa contoh peran dari seorang Data Engineer yang mungkin bersinggungan atau bahkan sama dengan peran Data Scientist dan Data Analyst !
2. Pengumpulan Data:
   * Data Engineer bertanggung jawab untuk mengumpulkan data dari berbagai sumber, membersihkannya, dan memastikan data tersebut tersedia dalam format yang sesuai untuk analisis.
   * Data Scientist juga bisa terlibat dalam pengumpulan data, terutama ketika data yang diperlukan tidak tersedia dan mereka perlu mencari atau menghasilkan data yang relevan.
3. Pengolahan Data:
   * Data Engineer melakukan pemrosesan data (ETL - Extract, Transform, Load) untuk mengubah data mentah menjadi format yang siap untuk analisis.
   * Data Scientist dapat melakukan transformasi data tambahan sesuai kebutuhan analisis mereka.
   * Data Analyst juga sering melakukan transformasi data untuk menjawab pertanyaan analisis mereka.
4. Perancangan Database:
   * Data Engineer merancang struktur database dan sistem penyimpanan data untuk mengoptimalkan penyimpanan dan akses data.
   * Data Scientist dan Data Analyst perlu memahami struktur data dan bagaimana mengaksesnya untuk mengerjakan analisis.
5. Pemeliharaan Data Pipeline:
   * Data Engineer memelihara data pipeline untuk memastikan data tersedia secara real-time atau sesuai jadwal.
   * Data Scientist dan Data Analyst bergantung pada data yang konsisten dan up-to-date untuk analisis mereka.
6. Keamanan Data:
   * Data Engineer harus memastikan bahwa data yang dikumpulkan dan disimpan aman dan sesuai dengan peraturan privasi.
   * Data Scientist dan Data Analyst harus mematuhi pedoman keamanan data saat melakukan analisis.
7. Jelaskan langkah-langkah proses ETL dan ELT yang berperan dalam pekerjaan seorang data engineer!

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

1. **Ekstraksi (Extract):**
   * Pada tahap ini, Data Engineer mengumpulkan data dari berbagai sumber, seperti database, file teks, API, atau sumber data lainnya.
   * Data dapat diekstrak dalam berbagai format, seperti CSV, JSON, XML, SQL, dan lain-lain.
2. **Transformasi (Transform):**
   * Setelah data diekstrak, Data Engineer melakukan serangkaian transformasi untuk membersihkan, mengubah, dan memproses data agar sesuai dengan kebutuhan analisis.
   * Transformasi ini mencakup pemilihan kolom, penggabungan data, pembersihan data yang salah, penghitungan agregat, pembuatan kolom turunan, dll.
3. **Pemuatan (Load):**
   * Data yang telah diubah melalui proses transformasi akan dimuat ke sistem penyimpanan data yang dituju. Ini bisa berupa basis data relasional, penyimpanan data berdistribusi, penyimpanan berbasis awan, atau tempat penyimpanan data lainnya.
   * Data Engineer memastikan bahwa data tersimpan dalam format yang cocok dengan struktur penyimpanan yang dipilih.

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

1. **Ekstraksi (Extract):**
   * Seperti pada proses ETL, tahap pertama adalah ekstraksi data dari berbagai sumber. Data Engineer mengumpulkan data dari sumber yang beragam, seperti database, file, API, dan lain-lain.
2. **Pemuatan (Load):**
   * Data yang diekstrak dimuat ke dalam sistem penyimpanan data yang dituju secepat mungkin. Ini bisa berupa data lake, data warehouse, atau penyimpanan data lainnya.
   * Data disimpan dalam bentuk mentah tanpa perlu transformasi awal.
3. **Transformasi (Transform):**
   * Transformasi data dilakukan setelah data dimuat ke dalam penyimpanan data. Data Engineer memproses data dalam sistem penyimpanan menggunakan bahasa pemrograman atau bahasa query, seperti SQL.
   * Transformasi ini mencakup penyatuan data, agregasi, pengolahan data, dan pembuatan tampilan data yang sesuai untuk analisis.