**FUNDAMENTAL OF DE - PART 1**

**TASK 1**

1. Peran utama seorang Data Engineer dalam ekosistem data adalah membangun dan mengelola ekosistem data yang berkualitas tinggi dan berintegritas, serta efektif yang mana ekosistem data inilah yang digunakan oleh organisasi / perusahaan dalam pengambilan keputusan sesuai dengan kebutuhan perusahaan atau *stakeholder*.

Berbeda dengan Data Analyst dan Data Scientist, dimana Data Analyst lebih mengutamakan keterampilan dalam mengolah, menginterpretasikan, dan menyajikan data dalam bentuk visualisasi data maupun laporan, sedangkan Data Scientist lebih bertanggung jawab dalam merumuskan hipotesis dan pembuatan model yang menggunakan teknik statistik dan *machine learning* dalam pemrosesan prediksi data, Data Engineer bertanggung jawab dalam merancang, membangun, dan memelihara infrastruktur data dengan memahami basis data, *data warehousing*, dan *tools* sedemikian hingga data dalam format yang siap digunakan oleh *data analyst*.

Beberapa peran utama dari seorang DE secara detail sebagai berikut.

1. Merancang dan Membangun Infrastruktur Data: Data engineer bertanggung jawab untuk merancang dan membangun infrastruktur data yang diperlukan untuk pengolahan dan penyimpanan data yang efisien. Ini meliputi pemilihan teknologi *database*, *data warehouse*, dan alat ETL yang sesuai dengan kebutuhan organisasi.
2. Mengembangkan Alur Kerja ETL ataupun ELT: Data engineer merancang, mengembangkan, dan mengelola alur kerja ETL (*Extract, Transform, Load*) ataupun ELT (*Extract, Load, Transform*) untuk mentransformasi data dari sumber ke penyimpanan data. Mereka memastikan alur kerja berjalan lancar, efisien, dan andal.
3. Memastikan Kualitas Data: Data engineer bertanggung jawab untuk memastikan kualitas data yang baik dengan melakukan pemrosesan data yang tepat, pemantauan kualitas data secara teratur, dan menangani masalah integritas data.
4. Mengoptimalkan Kinerja: Data engineer bekerja untuk mengoptimalkan kinerja infrastruktur data, alur kerja ETL, dan query data untuk memastikan data dapat diakses dan diproses dengan cepat dan efisien.
5. Kolaborasi dengan Tim Data: Data engineer bekerja sama dengan tim data lainnya, termasuk data analyst dan data scientist, untuk memahami kebutuhan analisis data dan memastikan infrastruktur data mendukung tujuan bisnis.
6. Keamanan Data: Data engineer memastikan data aman dari akses yang tidak sah atau kebocoran dengan menerapkan praktik keamanan data yang tepat, seperti enkripsi data dan akses kontrol.
7. Mengelola Metadata: Data engineer bertanggung jawab untuk mengelola metadata, termasuk informasi tentang asal-usul, struktur, dan pemilik data, untuk memudahkan pengelolaan dan pemahaman data.
8. Pemeliharaan dan Pemantauan Infrastruktur: Data engineer melakukan pemeliharaan rutin dan pemantauan infrastruktur data untuk memastikan keandalan, ketersediaan, dan kinerja sistem secara keseluruhan.
9. Peran seorang Data Engineer yang mungkin bersinggungan dengan Data Analyst ataupun Data Scientist adalah
10. Pemrosesan Data: Semua pelaku data memiliki peran yang hampir sama dalam pemrosesan data.

* Data Engineer memproses data dengan cara merancang dan mengembangkan alur kerja ETL ataupun ELT.
* Data Analyst memproses data dengan membersihkan data dan mengolah data yang ada.
* Data Scientist memproses data dengan membuat model dan *machine learning*.

1. Pemodelan Data:

* Data Engineer membuat model data untuk merancang struktur basis data atau *data warehouse* yang mendukung analisis data.
* Data Analyst membuat model analisis sederhana.
* Data Scientist membuat model data untuk menghasilkan prediksi data untuk kedepannya.

1. Memastikan Kualitas Data: Semua pelaku data memiliki peran yang sama dalam memastikan data yang berkualitas guna pengambilan keputusan yang tepat sesuai dengan kebutuhan organisasi / perusahaan, meskipun masing-masing peran memiliki tujuan yang berbeda menyesuaikan dengan tanggung jawab masing-masing, seperti

* Data Engineer yang memastikan kualitas data yang berintegritas dan berkualitas guna diproses sesuai dengan kebutuhan data analyst.
* Data Analyst memastikan kualitas data yang telah diolah memiliki interpretasi yang efektif guna pengambilan keputusan organisasi / perusahaan.

1. Kolaborasi dengan Tim Data: Semua pelaku data tentunya memiliki tanggung jawab yang sama dalam berkolaborasi dengan tim data yang lain supaya mampu mencapai tujuan bersama dalam suatu organisasi / perusahaan.
2. Langkah-langkah proses ETL dan ELT.

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

1. Ekstraksi (*Extract*):

* Mengidentifikasi asal / sumber data yang diperlukan.
* Mengambil data dari sumber dan mentransfernya ke dalam sistem ETL.

1. Transformasi (*Transform*):

* Membersihkan data dari sumber dan memastikan konsistensi.
* Mengubah format data agar sesuai dengan kebutuhan analisis.
* Menggabungkan data dari berbagai sumber.
* Menerapkan aturan bisnis atau transformasi kompleks lainnya.

1. Pemuatan (*Load*):

* Memuat data yang sudah ditransformasi ke dalam penyimpanan data, seperti *data warehouse* atau *data lake*.
* Memantau proses pemuatan untuk memastikan data dimuat dengan benar.

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

1. Ekstraksi (*Extract*):

* Mengidentifikasi asal / sumber data yang diperlukan.
* Mengambil data dari sumber dan langsung memuat data ke dalam penyimpanan data.

1. Pemuatan (*Load*):

* Memuat data yang sudah diekstraksi ke dalam penyimpanan data, tanpa transformasi yang signifikan.

1. Transformasi (*Transform*):

* Setelah data dimuat ke dalam penyimpanan data, melakukan transformasi data menggunakan alat atau bahasa pemrograman seperti SQL.
* Menerapkan aturan bisnis atau transformasi kompleks pada data yang sudah dimuat.