

## FUNDAMENTAL OF DE

### Task 1

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

Jawaban :

- Seorang Data Engineer dalam ekosistem data bertanggung jawab dalam merancang, membangun, dan memelihara infrastruktur dan system yang diperlukan untuk mengumpulkan, menyimpan data dalam skala besar sehingga data tersebut dapat digunakan untuk keperluan perusahaan.
- Seorang Data Engineer bertugas dalam menyiapkan data yang nanti akan diolah dan dianalisis oleh Data Analyst atau Data Scientist sehingga dari data tersebut didapat insight, ataupun model machine learning yang dibutuhkan. Data Analyst dan Data Scientist membutuhkan data yang bersih dan berkualitas, dimana data tersebut disediakan oleh Seorang Data Engineer.

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

Jawaban :

- Peran yang Data Engineer yang kemungkinan bersinggungan dengan Data Analyst dan Data Scientist diantaranya yang berkaitan dengan pemrosesan data diantaranya data formatting, data cleaning, handling missing values, bahkan scaling.

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

Jawaban :

❖ ETL:

- Langkah pertama dalam ETL adalah Extract atau mengextraksi data dari berbagai sumber berupa database, file, API, dan dari sumberlainya.
- Setelah data diekstraksi, langkah berikutnya adalah **mentransformasi** data menjadi format yang sesuai dan berguna untuk analisis lebih lanjut. Transformasi ini mungkin melibatkan pembersihan data, pemfilteran, normalisasi, penggabungan data dari beberapa sumber, dan perhitungan agregat.
- Setelah data telah diolah dan ditransformasi, langkah terakhir adalah memuatnya ke sistem penyimpanan data yang dituju, seperti data warehouse atau data lake.

❖ ELT :

- Seperti dalam ETL, langkah pertama adalah mengekstraksi data dari berbagai sumber.
- Setelah data diekstraksi, langkah berikutnya adalah memuatnya ke sistem penyimpanan data tanpa transformasi yang signifikan.
- Setelah data dimuat ke sistem penyimpanan data, transformasi dilakukan di dalam sistem penyimpanan data itu sendiri.