

## FUNDAMENTAL OF DE

### PART 1 - Fundamental DE

#### TASK 1

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

Jawab:

- Peran utama Data Engineering adalah proses mengambil data mentah dan membersihkannya, mengubahnya (mentransformasikannya), menganalisanya dan memublikasikannya agar bermakna dan berguna untuk orang-orang dan organisasi.
- Peran data engineer lebih fokus membangun dan mengelola infrastruktur data.
- Sedangkan data scientist mengembangkan model prediktif dan analitik lanjutan menggunakan pembelajaran mesin dan teknik statistik.
- Dan data analyst menganalisis data untuk menghasilkan wawasan dan laporan yang mendukung keputusan bisnis.

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

Jawab:

Peran data engineer yang bersinggungan atau bahkan sama dengan peran Data Scientist dan Data Analyst adalah data cleaning, anomaly detection, dan data preparation.

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

Jawab:

Langkah-langkah proses ETL

- a. Pipeline data mengekstrak data mentah dari berbagai sumber
- b. Menyimpannya di lokasi sementara yang disebut staging area
- c. Mengubah data di staging area
- d. Memuatnya ke dalam data lake atau data warehouse

Langkah-langkah proses ELT

- a. Pipeline data mengekstrak dan memuat data mentah langsung ke dalam data lake atau data warehouse. Data lake biasanya digunakan untuk penyimpanan data

mentah dalam berbagai format, sementara data warehouse digunakan untuk penyimpanan data terstruktur yang siap dianalisis.

- b. Kemudian melakukan transformasi/perubahan setelah memindahkan informasi ke data warehouse atau data lake. Transformasi ini dapat mencakup pembersihan data, agregasi, pengindeksan, dan konversi format data.