

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

Proses ETL (Extract, Transform, Load) dan ELT (Extract, Load, Transform)

ETL (Extract, Transform, Load)

1. Extract (Ekstraksi)

- **Deskripsi:** Mengambil data dari berbagai sumber (database, file, API, dll.).
- **Peran Data Engineer:** Menulis skrip atau menggunakan alat ETL untuk mengakses dan mengumpulkan data secara efisien.

2. Transform (Transformasi)

- **Deskripsi:** Mengubah data sesuai kebutuhan analisis (pembersihan, agregasi, normalisasi, dll.).
- **Peran Data Engineer:** Merancang dan mengimplementasikan logika transformasi dengan alat ETL atau kode dalam Python/SQL.

3. Load (Pemuatan)

- **Deskripsi:** Memuat data yang sudah ditransformasi ke sistem penyimpanan akhir (data warehouse).
- **Peran Data Engineer:** Memastikan data dimuat dengan benar dan melakukan optimisasi performa serta integritas data.

ELT (Extract, Load, Transform)

1. Extract (Ekstraksi)

- **Deskripsi:** Sama seperti ETL, mengambil data dari berbagai sumber.
- **Peran Data Engineer:** Menghubungkan ke sumber data dan mengumpulkan data yang diperlukan.

2. Load (Pemuatan)

- **Deskripsi:** Data mentah langsung dimuat ke sistem penyimpanan target (data lake atau data warehouse).
- **Peran Data Engineer:** Mengelola proses pemuatan data mentah agar cepat tersedia dan dapat diakses.

3. Transform (Transformasi)

- **Deskripsi:** Transformasi data dilakukan di dalam sistem penyimpanan menggunakan kapasitas pemrosesan yang ada.
- **Peran Data Engineer:** Menulis query SQL atau menggunakan alat transformasi data dalam sistem penyimpanan untuk mengubah data sesuai kebutuhan.

Perbandingan ETL dan ELT

- **ETL:** Cocok untuk transformasi data yang kompleks sebelum pemuatan; sering digunakan dengan data warehouse tradisional.
- **ELT:** Efisien untuk memuat dan memproses data dalam jumlah besar dengan cepat; memanfaatkan kekuatan pemrosesan data warehouse modern seperti Google BigQuery atau Amazon Redshift.

Contoh Implementasi

- **ETL:** Menggunakan Apache NiFi atau Talend untuk mengambil data dari database, mengubahnya, dan memuatnya ke data warehouse seperti Amazon Redshift.
- **ELT:** Menggunakan Apache Kafka untuk mengumpulkan data mentah dan memuatnya ke data lake seperti Amazon S3, lalu menggunakan AWS Glue atau query SQL di Amazon Redshift untuk transformasi data.

Kedua pendekatan ini membantu Data Engineer memastikan data berkualitas tinggi dan siap digunakan untuk analisis dan pengambilan keputusan bisnis.