

Nama : Karno

Jurusan : DE_01

1. Seorang Data Engineer memiliki peran yang sangat penting dalam ekosistem data. Peran utama seorang Data Engineer adalah merancang, mengembangkan, dan memelihara infrastruktur teknis yang diperlukan untuk mengumpulkan, menyimpan, dan mengelola data. Berikut ini beberapa tugas dan tanggung jawab utama seorang Data Engineer:
 - a. Pengumpulan Data: Data Engineer bertanggung jawab untuk mengumpulkan data dari berbagai sumber, termasuk basis data, sistem eksternal, streaming data, dan lainnya. Mereka harus memastikan data-data ini dikumpulkan dengan benar dan efisien.
 - b. ETL (Extract, Transform, Load): Data Engineer merancang dan membangun alur kerja ETL untuk mengambil data mentah, mengubahnya sesuai kebutuhan, dan memuatnya ke dalam sistem penyimpanan data.
 - c. Data Warehousing: Data Engineer sering kali bekerja dengan sistem penyimpanan data seperti data warehouse atau data lake untuk menyimpan data secara efisien dan memastikan aksesibilitas yang baik.
 - d. Perancangan Database: Mereka merancang skema database dan tabel yang sesuai dengan kebutuhan organisasi.
 - e. Pemeliharaan Infrastruktur: Data Engineer memelihara infrastruktur teknis yang diperlukan untuk pengelolaan data, termasuk server, basis data, alat-alat ETL, dan perangkat keras lainnya.
 - f. Pengamanan Data: Mereka harus memastikan keamanan data dengan menerapkan langkah-langkah perlindungan data yang sesuai.
 - g. Skalabilitas: Data Engineer juga harus memastikan infrastruktur dapat dengan mudah diperluas untuk menangani pertumbuhan data yang signifikan.

Peran Data Engineer berbeda dengan Data Scientist dan Data Analyst dalam hal fokus dan tanggung jawab:

1. Data Scientist: Data Scientist bertanggung jawab untuk menggali wawasan dari data, membangun model prediksi, melakukan analisis statistik, dan menghasilkan wawasan yang dapat membantu organisasi membuat keputusan. Mereka sering menggunakan bahasa pemrograman seperti Python atau R untuk analisis data.
2. Data Analyst: Data Analyst, di sisi lain, lebih fokus pada pengambilan data yang sudah ada, membersihkan data, dan menyajikannya dalam bentuk laporan, grafik, atau visualisasi untuk membantu pemangku kepentingan dalam pengambilan keputusan. Mereka biasanya menggunakan alat analisis data seperti Excel, SQL, atau alat visualisasi data.

Sementara Data Engineer lebih berfokus pada aspek teknis infrastruktur data, Data Scientist dan Data Analyst lebih berfokus pada analisis data dan interpretasi wawasan. Meskipun peran-peran ini memiliki perbedaan yang signifikan dalam tanggung jawab dan fokus, mereka sering bekerja sama dalam ekosistem data untuk menciptakan nilai dari data organisasi.

2. Beberapa peran dari seorang Data Engineer mungkin bersinggungan atau bahkan sama dengan peran Data Scientist dan Data Analyst :

- a. Pemahaman Bisnis: Data Engineer, Data Scientist, dan Data Analyst, perlu memiliki pemahaman mendalam tentang bisnis dan tujuan organisasi. Mereka harus tahu apa yang ingin dicapai organisasi melalui pengolahan data.
- b. Pengolahan Data Dasar: Meskipun Data Engineer bertanggung jawab untuk ETL dan pengolahan data mentah, Data Scientist dan Data Analyst juga mungkin perlu melakukan pengolahan data dasar untuk persiapan analisis atau pemodelan.
- c. Penyimpanan dan Manajemen Data: Data Scientist dan Data Analyst dapat memerlukan pengetahuan dasar dalam penyimpanan dan manajemen data untuk mengakses data dengan benar. Data Engineer adalah yang merancang dan mengelola infrastruktur data ini.
- d. Penyelidikan Data: Data Scientist dan Data Analyst sering harus melakukan penyelidikan data untuk memahami aspek data yang lebih baik sebelum melakukan analisis atau pembuatan laporan.
- e. Penggunaan Alat Analisis Data: Data Engineer mungkin perlu menggunakan alat analisis data dalam beberapa situasi, terutama ketika mereka harus memeriksa data selama pengembangan dan pemeliharaan infrastruktur data.
- f. Pemahaman Algoritma: Meskipun Data Scientist adalah yang paling terlibat dalam pengembangan algoritma dan model prediksi, Data Analyst dan Data Engineer juga mungkin perlu memahami dasar-dasar algoritma data untuk memahami hasil analisis.
- g. Visualisasi Data: Data Engineer, Data Scientist, dan Data Analyst mungkin perlu menciptakan visualisasi data untuk menyajikan hasil analisis atau mengkomunikasikan wawasan kepada pemangku kepentingan.

3. Langkah-langkah ETL dan ELT :

ETL (ekstraksi Transform Load)

1. Ekstraksi (Extract):

- Pengumpulan data dari berbagai sumber seperti database, file teks, API, atau sistem eksternal lainnya.
- Data biasanya diekstrak dalam bentuk mentah (raw data) dan disimpan dalam zona ekstraksi.

2. Transformasi (Transform):

- Proses pembersihan, pemrosesan, dan transformasi data mentah menjadi format yang lebih sesuai untuk analisis atau aplikasi.
- Transformasi dapat mencakup penghapusan data yang tidak relevan, penggabungan data dari berbagai sumber, perhitungan, normalisasi, dan pemformatan data.

- Proses ETL ini sering memanfaatkan alat seperti Apache Spark, Apache Flink, atau Python untuk pemrosesan data.

3. Pemuatan (Load):

- Data yang telah diolah kemudian dimuat ke dalam warehouse data atau penyimpanan data lainnya.
- Warehouse data adalah basis data yang dirancang khusus untuk analisis dan laporan, seperti Amazon Redshift, Google BigQuery, atau Snowflake.

ELT (Ekstraksi Load Transform)

1. Ekstraksi (Extract):

- Mirip dengan langkah pertama dalam ETL, data diambil dari berbagai sumber dan dimuat ke dalam penyimpanan data, seperti data lake atau data warehouse.

2. Pemuatan (Load):

- Data dimuat ke dalam penyimpanan data tanpa transformasi yang signifikan.
- Data ini dimuat dalam bentuk mentah (raw data) ke penyimpanan data yang mendukung struktur fleksibel seperti data lake.

3. Transformasi (Transform):

- Setelah data ada di penyimpanan data, transformasi dilakukan menggunakan alat analisis data seperti SQL, Apache Spark, atau Python.
- Proses transformasi ini dapat mencakup pembuatan tabel, perhitungan agregat, join, dan pemrosesan data lebih lanjut sesuai kebutuhan analisis.

Perbedaan utama antara ETL dan ELT adalah urutan transformasi. Dalam ETL, transformasi dilakukan sebelum data dimuat ke penyimpanan data, sementara dalam ELT, transformasi dilakukan setelah data dimuat ke penyimpanan data.