

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

Seorang Data Engineer bertanggung jawab untuk mempersiapkan data secara efisien yang nantinya dapat diolah oleh Data Scientist dan Data Analyst, seorang Data Engineer bertugas untuk mendesain dan membangun sebuah infrastruktur yang dapat menampung, mengolah dan memproses data serta memelihara/memonitor infrastruktur tersebut.

Adapun perbedaan dengan Data Scientist dan Data Analyst adalah:

Data Scientist bertanggung jawab mengembangkan sebuah model prediktif dan algoritma untuk mengidentifikasi tren, pola, dan wawasan dari sebuah kumpulan data.

Data Analyst berfokus untuk menganalisa sebuah kumpulan data historis untuk memberikan insight berupa tren, pola dan informasi yang berguna bagi pengambilan keputusan bisnis

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

Data Engineer, Data Scientist dan Data Analyst memiliki kesamaan dalam beberapa hal seperti ketiga role tersebut terkadang memiliki peran dalam melakukan cleaning dan preparation data . Di beberapa perusahaan Data Engineer juga dituntut untuk menganalisa data untuk memberikan insight bisnis sehingga hal ini sama dengan role dari Data Analyst.

Selain itu dalam penerapan industri, ketiga role ini akan berkerja sama dalam sebuah tim yang terintegrasi, mulai dari pengumpulan data, analisa data hingga melakukan forecasting dari data yang telah dianalisa.

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

ETL (Extract-Transform-Load)

Proses ini dimulai dengan meng-extract raw data yang dapat berupa database, JSON, CSV, API, Excel, dll dari berbagai sumber dan dikumpulkan ke dalam satu wadah tertentu. Kemudian raw data akan langsung di-transform dalam beberapa tahap yaitu pembersihan data, standarisasi data dan pengelompokan data. Raw data yang telah di-transform akan di-load ke data warehouse sehingga menjadi data yang siap digunakan

ELT (Extract-Load-Transform)

Proses ini dimulai dengan meng-extract raw data yang dapat berupa database, JSON, CSV,

API, Excel, dll dari berbagai sumber dan langsung di-load ke dalam sebuah data warehouse. Setelah data di-load baru lah raw data tersebut akan di-transform dalam beberapa tahap yaitu pembersihan data, standarisasi data dan pengelompokan data. Setelah data selesai di-transform, data yang bersih atau siap digunakan akan di-load kembali ke data warehouse semula,