

# DATA ENGINEER

## UNIT 2 – PART 1 – Fundamental Data Engineer

Presentation by  
Nurul Kholifah

# BAGIAN 1

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

Dalam ekosistem data, data engineer memiliki peran utama mengumpulkan data dari berbagai sumber, termasuk database, sistem eksternal, API, dan sumber data lainnya, mentransformasikannya menjadi format yang sesuai serta memastikan data dikumpulkan dengan cara yang efisien sehingga dapat diakses oleh Data Scientist dan Data Analyst.

Berbeda dengan Data Engineer, Data Scientist fokus pada analisis data, membersihkan data dari noise atau kesalahan, eksplorasi data untuk menemukan wawasan, membangun model, menguji model, dan mengimplementasikan solusi yang dihasilkan untuk selanjutnya membantu perusahaan merancang strategi bisnis, mengidentifikasi peluang baru, dan mengembangkan produk atau layanan yang lebih baik.

Sedangkan Data Analyst menganalisis data untuk memberikan wawasan yang bermanfaat dalam pengambilan keputusan sehari-hari, menghasilkan laporan, grafik, dan visualisasi data yang dapat dipahami oleh orang lain.

## BAGIAN 2

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

Berikut adalah beberapa contoh peran Data Engineer yang bersinggungan dengan Data Scientist dan Data Analyst

1. Data Preparation. Peran Data Engineer mengumpulkan data dari berbagai sumber membutuhkan koordinasi dengan Data Scientist dan Data Analyst untuk memastikan data tersedia dalam format yang sesuai dengan kebutuhan sehingga dapat digunakan oleh Data Scientist dan Data Analyst. Ketiganya perlu mengetahui dengan baik sumber data dan struktur data.
2. Programming dan Tools. Data Engineer, Data Scientist, dan Data Analyst biasanya menggunakan bahasa pemrograman dan alat analisis data yang sama seperti Python, R, dan SQL.

## BAGIAN 3

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

### 01 ETL (Extract Transform Load)

#### Extract

Langkah ini melibatkan ekstraksi data dari berbagai sumber seperti database, file teks, API, dan lainnya. Data engineer harus memahami sumber data dan menggunakan alat ekstraksi yang sesuai, seperti ETL tools, untuk mengambil data.

#### Transform

Transformasi melibatkan pembersihan, penyederhanaan, penyatuan, dan perubahan format data. Data engineer menggunakan alat seperti SQL, Python, atau alat ETL untuk mengubah data menjadi bentuk yang berguna dan siap untuk analisis.

#### Load

Data warehouse atau penyimpanan data lainnya, yang biasanya memiliki struktur yang telah ditentukan. Proses pemuatan ini memastikan data tersedia untuk analisis.

### 02 ELT (Extract Load Transform)

#### Extract

Data biasanya diekstraksi dari sumber ke dalam penyimpanan data (data warehouse) sebelum transformasi. Proses ini dapat melibatkan ekstraksi penuh atau sebagian data tanpa banyak perubahan struktur.

#### Load

Data biasanya dimuat terlebih dahulu ke dalam penyimpanan data tanpa banyak transformasi. Transformasi berikutnya dilakukan pada data yang ada dalam penyimpanan data, dan hasilnya disimpan dalam penyimpanan data yang sama atau berbeda.

#### Transform

Transformasi dalam ELT biasanya terjadi setelah data dimuat ke dalam penyimpanan data. Proses ini melibatkan penggunaan alat seperti Apache Spark, Hadoop, atau alat lain yang dapat mengolah data dalam skala besar. Transformasi dapat menjadi tugas yang sangat komputasi-berat.

The background is a dark blue field decorated with various geometric elements. It includes numerous small squares in white, light blue, and light orange, some of which are solid while others are outlines. Additionally, there are several thin, vertical white lines of varying lengths scattered across the composition. The central text 'TERIMA KASIH' is prominently displayed in a clean, sans-serif font, with 'TERIMA' in white and 'KASIH' in a vibrant light blue.

TERIMA KASIH