

1. Peran utama seorang Data Engineer dalam ekosistem adalah dimulai dari mengumpulkan data, mengekstraksi data, melakukan transformasi dan cleansing, menganalisis, hingga mempublikasikan hasil data yang diolah.

Yang membedakan dengan Data Analyst maupun Data Scientist adalah Data Engineer tidak sampai pada tahap menemukan insight, melakukan forecast/predict, maupun membuat decision making.

2. Contoh peran seorang Data Engineer yang bersinggungan dengan Data Analyst maupun Data Scientist adalah mengumpulkan (collect) data, menyimpan (store) data, cleansing data, dan analisis data serta menggunakan tools yang sama.

3. Proses ETL:

- a. Mengekstrak raw data dari beberapa data source
- b. Menyimpan data sementara di staging area
- c. Transform data dari staging area
- d. Load data dan menyimpannya di data lake/warehouse

Proses ELT:

- a. Mengekstrak raw data dan langsung load data secara langsung ke data lake
- b. Melakukan transform/change data setelah memindahkan informasi ke data warehouse