**NAMA : TARISA DINA LUTFIANA**

· Task

Soal Prioritas 1

1. **Sebutkan dan jelaskan berbagai skill yang dibutuhkan untuk menjadi seorang data engineer! (minimal 3)**

·Memiliki skill Python yang merupakan bahasa pemrograman mudah dipelajari, dan dapat membuat solusi dan memperluas sesuai perkembangan data.

·Memiliki skill SQL(Structured Query Language) yang dapat memudahkan dalam pembuatan query dalam database yang akan dilakukan seorang data engineer.

·Memiliki skill dalam Cloud (AWS, GCP) yang digunakan dalam pelayanan data engineering.

2. **Sebutkan dan jelaskan urgensi / pentingnya peran data engineer dalam industri di saat ini!**

Pada pentingnya peran data engineer dalam industri saat ini yang pertama dalam pengambilan keputusan. Data engineer dapat berfokus pada pengaturan dalam pengisian database untuk mengembangkan skema table database. Tugas utama dalam data engineering yaitu dengan dapat memproses data pada sumber penyimpanan yang sesuai, contohnya seperti data warehouse. Pada data engineer dapat mempersiapkan data yang siap digunakan dengan contoh seperti reliable data flow, infrastructure, pipelines, dan lain sebagainya, untuk data yang disiapkan dapat digunakan oleh data analyst.

Soal Prioritas 2

**1.** **Sebuah rumah sakit bernama Alta Medika ingin mengelola berbagai data yang telah dikumpulkan seperti data pasien, data komplain, data survei kepuasan pelayanan, data dokter dan data lainnya. Untuk mengelola berbagai data tersebut dapat dilakukan dengan mengembangkan sebuah data pipeline. Berdasarkan kasus ini sebutkan jenis data pipeline yang cocok digunakan beserta alasannya!**

Pada kasus diatas jenis data pipeline yang digunakan yaitu :

· Batch Data Pipeline yang merupakan cocok dalam kasus rumah sakit Alta Medika, karena pada proses pengolahan data ini dapat dilakukan secara periodik, dan data yang sudah dikumpulkan dengan waktu tertentu, seperti hitungan menit, jam , dan bahkan hari. Serta diproses secara periodic query sesuai yang ditentukan, selanjutnya pada hasilnya dapat ditulis di database.

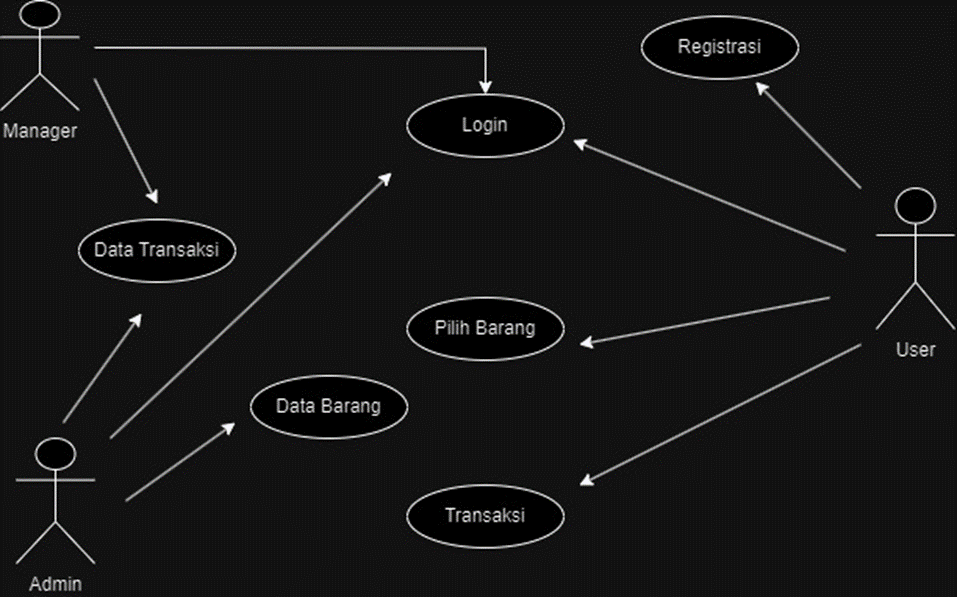
· Stream Data Pipeline yang termasuk cocok dalam kasus diatas, karena pada proses ini berbeda dengan batch data pipeline yang menggunakan data yang sudah dikumpulkan, sedangkan pada stream data pipeline ini menggunakan data real time atau mendapatkan data yang cepat masuk yang berfungsi untuk data complain atau kepuasan pelayanan. Streaming data merupakan proses untuk dibaca pada sekumpulan data secara continuous.

Soal Eksplorasi

**1.** **Sebuah platform e-commerce ingin mendapatkan pengetahuan mengenai jumlah transaksi per bulan, jenis barang yang sering dibeli serta persentase keuntungan yang diperoleh. Untuk mendapatkan pengetahuan tersebut dapat dilakukan dengan mengembangkan sebuah data pipeline. Berdasarkan kasus ini sebutkan jenis data pipeline yang cocok digunakan beserta alasannya!**

Batch Data Pipeline yang merupakan cocok pada kasus platform e-commerce diatas yang membutuhkan informasi jumlah data transaksi perbulan, jenis barang yang dibeli, serta keuntungan. Pada data yang sudah dikumpulkan dilakukan analisis secara pada periode tretentu, misalnya pada akhir bulan.

**2.** **Berdasarkan soal nomor 1, buat diagram sederhana dari data pipeline yang dipilih. Diagram dapat dibuat dengan draw.io, excalidraw atau tools lainnya yang sejenis.**

****