**Task 1**

Soal Prioritas 1

1. Menurut website Digital Skola, ada 5 skill yang dibutuhkan untuk menjadi seorang data engineer, diantaranya yaitu :

* SQL (*Structured Query Language*), bahasa pemrograman yang dimanfaatkan untuk berinteraksi dengan sistem basis data.
* Teknik Data Modeling, keterampilan yang harus dimiliki oleh seorang data engineer untuk merancang model data guna menentukan informasi yang dibutuhkan.
* Bahasa Pemrograman Python menjadi salah satu bahasa pemrograman yang sering digunakan oleh data engineer.
* Hadoop, alat yang umum digunakan dalam pengolahan big data
* Layanan AWS Cloud populer digunakan karena menyediakan berbagai layanan yang mendukung peran seorang data engineer.

1. Menurut DQlab, urgensi peran seorang data engineer dalam industri saat ini dapat dibedakan berdasarkan ukuran tim, yang terdiri dari tim kecil, menengah, dan besar.

* Dalam tim kecil, seorang data engineer memiliki tanggung jawab menyeluruh terhadap semua langkah dalam pemrosesan data.
* Dalam tim menengah, peran data engineer lebih difokuskan pada pengembangan dan pemeliharaan pipa data (pipeline).
* Dalam tim besar, seorang data engineer berperan sebagai pusat pengelolaan database, termasuk pengaturan, pengisian, dan pengembangan skema tabel di dalam database.

Sumber referensi

1. <https://digitalskola.com/blog/data-engineer/data-engineer-skills>
2. <https://dqlab.id/pahami-role-data-engineer-bersama-dqlab>

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 data tersebut secara terjadwal, batch data pipeline dapat digunakan karena memungkinkan pengolahan data dalam jumlah besar yang telah dikumpulkan secara efisien.

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, jika ingin mengelola data secara terjadwal dalam jumlah besar, maka dapat menggunakan batch data pipeline. Batch data pipeline cocok digunakan karena data transaksi biasanya tersedia dalam jumlah besar dapat diolah secara efisien dalam batch.
2. Diagram batch data pipeline

