**Part 1**

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

* **Peran Utama Data Engineer**

1. Desain dan Pengembangan Infrastruktur Data:

* Membuat dan memelihara arsitektur data yang mendukung pengumpulan, pengolahan, dan penyimpanan data dari berbagai sumber.
* Mengembangkan pipeline ETL (Extract, Transform, Load) untuk memproses data mentah menjadi format yang dapat digunakan oleh pengguna akhir.

1. Manajemen dan Optimasi Data:

* Mengelola database dan sistem penyimpanan data, memastikan kinerja yang optimal dan skalabilitas.
* Mengimplementasikan solusi untuk menangani big data dengan menggunakan teknologi seperti Hadoop, Spark, dan Kafka.

1. Keamanan dan Kepatuhan Data:

* Menerapkan langkah-langkah keamanan untuk melindungi data sensitif dan memastikan kepatuhan terhadap regulasi dan standar industri.
* Memantau dan mengelola akses ke data untuk memastikan integritas dan keamanan data.

### **Perbedaan dengan Data Scientist dan Data Analyst**

1. **Data Engineer:**

* Fokus pada pembangunan dan pengelolaan infrastruktur data serta pipeline untuk pengumpulan, penyimpanan, dan pemrosesan data.
* Lebih terfokus pada keahlian teknis seperti pemrograman (Python, Java, Scala), manajemen database (SQL, NoSQL), dan teknologi big data.
* Membangun dan mengoptimalkan sistem data yang andal dan efisien.

1. **Data Scientist:**

* Fokus pada analisis data, membangun model prediktif, dan menggunakan teknik machine learning untuk mengekstrak wawasan dari data.
* lebih terfokus pada statistik, analisis data, dan teknik machine learning.
* Menghasilkan model analitik dan prediktif serta wawasan yang dapat digunakan untuk pengambilan keputusan bisnis.

1. **Data Analyst:**

* Fokus pada analisis data untuk menghasilkan laporan, visualisasi, dan wawasan bisnis. Mereka menggunakan data yang sudah terstruktur dan disiapkan oleh data engineer.
* membuat laporan, dashboard, dan analisis data yang membantu bisnis dalam pengambilan keputusan.

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

Contoh Peran Bersinggungan antara Data Engineer dan Data Scientist:

1. **Pemrosesan Data:**
   * **Data Engineer:** Menyusun pipeline data untuk mengumpulkan, membersihkan, dan mengubah data menjadi format yang siap untuk analisis.
   * **Data Scientist:** Menggunakan data yang sudah diproses oleh Data Engineer.
2. **Eksplorasi Data:**
   * **Data Engineer:** Membantu menyediakan data dalam bentuk yang bisa diakses oleh Data Scientist untuk eksplorasi awal.
   * **Data Scientist:** Melakukan eksplorasi data untuk memahami pola, distribusi, dan anomali dalam data.

### Contoh Peran Bersinggungan antara Data Engineer dan Data Analyst:

1. **Pengelolaan Data:**
   * **Data Engineer:** Mengelola database dan memastikan data yang tersedia dapat diakses dengan efisien oleh pengguna lain.
   * **Data Analyst:** Mengambil dan menganalisis data dari database yang dikelola oleh Data Engineer untuk membuat laporan dan visualisasi.
2. **Pembuatan Kueri:**
   * **Data Engineer:** Membuat dan mengoptimalkan kueri yang kompleks untuk ekstraksi data.
   * **Data Analyst:** Menulis kueri SQL untuk mengakses dan menganalisis data, kadang-kadang memanfaatkan kueri yang telah dioptimalkan oleh Data Engineer.
3. Jelaskan langkah-langkah proses ETL dan ELT yang berperan dalam pekerjaan seorang data engineer !

### **Proses ETL (Extract, Transform, Load)**

1. **Extract (Ekstraksi):**
   * **Tujuan:** Mengambil data dari berbagai sumber yang berbeda seperti database relasional, file flat (CSV, JSON, XML), API, dan sumber lainnya.
   * **Tugas Data Engineer:** Menulis skrip atau menggunakan alat ETL untuk mengakses dan mengekstrak data dari sumber-sumber tersebut.
2. **Transform (Transformasi):**
   * **Tujuan:** Membersihkan, memperkaya, dan mengubah data menjadi format yang sesuai untuk analisis.
   * **Tugas Data Engineer:**
     + Melakukan operasi seperti penghapusan duplikasi, penanganan nilai yang hilang, normalisasi, agregasi, dan penggabungan data dari berbagai sumber.
     + Mengimplementasikan aturan bisnis dan logika transformasi yang diperlukan untuk memastikan data berkualitas tinggi.
3. **Load (Pemuatan):**
   * **Tujuan:** Memuat data yang telah diproses ke dalam data warehouse atau sistem penyimpanan tujuan lainnya.
   * **Tugas Data Engineer:**
     + Menentukan metode pemuatan yang efisien (batch loading atau incremental loading) tergantung pada kebutuhan sistem.
     + Mengoptimalkan proses pemuatan untuk memastikan kinerja yang baik dan minimal gangguan pada sistem tujuan.

### **Proses ELT (Extract, Load, Transform)**

1. **Extract (Ekstraksi):**
   * **Tujuan:** Mengambil data dari berbagai sumber seperti pada proses ETL.
   * **Tugas Data Engineer:** Sama seperti dalam ETL, mengekstrak data dari berbagai sumber dengan menulis skrip atau menggunakan alat ETL.
2. **Load (Pemuatan):**
   * **Tujuan:** Memuat data mentah langsung ke dalam data lake atau data warehouse sebelum dilakukan transformasi.
   * **Tugas Data Engineer:**
     + Menggunakan alat dan teknologi yang mendukung pemuatan data besar secara efisien ke dalam data lake atau data warehouse (misalnya, Amazon S3, Google Cloud Storage, Hadoop HDFS).
     + Memastikan data mentah tersedia untuk transformasi dan analisis selanjutnya.
3. **Transform (Transformasi):**
   * **Tujuan:** Mengubah data yang sudah dimuat dalam data lake atau data warehouse.
   * **Tugas Data Engineer:** Menulis kueri SQL atau menggunakan alat big data (misalnya, Apache Spark, Databricks) untuk melakukan transformasi data.