



1. Cari Studi Kasus Pengelolaan Data

- Cari studi kasus pengelolaan data dalam konteks perusahaan atau organisasi pada sumber data publik online, misalnya Kaggle, data.gov, atau repositori data dummy.
- Pilih minimal 4 tabel sebagai berikut, sebagai contoh:
 - **Data Karyawan** – atribut seperti ID karyawan, nama, jabatan, departemen, gaji, tanggal masuk, dll.
 - **Data Penjualan** – atribut seperti ID transaksi, tanggal, produk, jumlah, harga, pelanggan, dll.
 - **Data Inventaris** – atribut seperti ID barang, nama barang, kategori, jumlah stok, lokasi gudang, dll.
 - **Data Pelanggan** – atribut seperti ID pelanggan, nama, alamat, histori pembelian, kategori pelanggan, dll.
- Pastikan data dummy yang dipilih tersedia, lengkap, dan dapat digunakan untuk pembangunan proyek ETL dan data warehouse.

2. Pilih Studi Kasus yang Memenuhi Kriteria

- Pilih satu studi kasus yang paling lengkap dan sesuai untuk kebutuhan analisis ETL dan desain data warehouse.
- Pastikan data mudah diakses dan digunakan secara legal untuk proyek.

3. Ambil Data Dummy dari Studi Kasus Terpilih

- Download atau akses data dummy tersebut.
- Periksa format data (CSV, JSON, Excel, dsb) dan atribut yang tersedia.

4. Rancang Skema Bintang (Star Schema) untuk Data Warehouse

- Tentukan tabel fakta dan tabel dimensi berdasarkan data studi kasus.
- Contoh untuk kasus data karyawan:
 - Tabel Fakta: Fact_Penggajian (atribut: gaji, bonus, tanggal pembayaran, id_karyawan, id_departemen, dll)
 - Tabel Dimensi:
 - Dim_Karyawan (id_karyawan, nama, tanggal_lahir, jenis_kelamin, status_kerja, dll)
 - Dim_Departemen (id_departemen, nama_departemen, lokasi)
 - Dim_Jabatan (id_jabatan, nama_jabatan, level)
 - Dim_Waktu (id_waktu, tanggal, bulan, tahun)

- Untuk kasus lain, sesuaikan skema, misal:
 - Data Penjualan: Fact_Penjualan, Dim_Produk, Dim_Pelanggan, Dim_Waktu
 - Data Inventaris: Fact_Stok, Dim_Barang, Dim_Gudang, Dim_Waktu
- Tentukan relasi antar tabel melalui foreign key.

5. Rancang Proyek ETL Berdasarkan Skema Bintang

- Buat pipeline ETL yang terdiri dari:
 - **Extract:** Mengambil data dari sumber data dummy (file CSV, database, API, dsb).
 - **Transform:** Membersihkan data, mengubah format data, konversi tipe data, agregasi, dan transformasi lainnya sesuai kebutuhan.
 - **Load:** Memasukkan data yang sudah bersih dan siap ke dalam tabel fakta dan dimensi di data warehouse.

6. Jelaskan Langkah-Langkah Proses ETL

- **Extract:** Membaca data dummy dari sumber, misalnya membaca file CSV berisi data karyawan atau data lain.
- **Transform:** Melakukan pembersihan data seperti memperbaiki format tanggal, normalisasi nama atau kategori, menghapus data duplikat, menyesuaikan struktur data sesuai skema star schema.
- **Load:** Memasukkan data ke tabel fakta dan dimensi dengan memperhatikan integritas data, kunci primer, dan relasi antar tabel.

7. Buat Laporan/Dokumentasi Proyek ETL

- Jelaskan seluruh proses pembangunan proyek ETL mulai dari pengambilan data dummy, perancangan skema, implementasi pipeline ETL.
- Jelaskan hasil yang diperoleh, misalnya data yang sudah terstruktur di data warehouse dan contoh hasil query , tampilkan dengan visualisasi dashboard.
- Analisis KPI (Key Performance Indicators), contoh:
 1. Total penjualan per periode tertentu
 2. Top 5 Produk/Menu Terlaris

8. Pengumpulan: Kumpulkan semua file proyek di repository GitHub kelas masing-masing, dengan **folder terpisah per individu**. Sertakan **file dengan ekstensi .etl** yang berisi pipeline ETL dan **laporan dalam format PDF** sebagai dokumentasi. (dikerjakan secara kelompok)

--- GOOD LUCK ---