**Flujo del Pipeline ETL para Análisis de Ventas**

**Descripción del Flujo**

El pipeline ETL (Extract, Transform, Load) sigue estos pasos:

1. **Extracción**: Se leen los archivos CSV (products.csv, customers.csv, orders.csv, order\_details.csv) desde la carpeta C:/Users/El amor de tu vida/OneDrive/ITLA/2025 c3/BIG DATA/etl/ventas/data/Archivo CSV Análisis de Ventas-20250924/.
2. **Transformación**:
   * Se limpian valores nulos y duplicados.
   * Se normalizan formatos (por ejemplo, Price como decimal, OrderDate como fecha, Phone eliminando caracteres no deseados).
   * Se integra un ejemplo de datos externos (API ficticia) para actualizar precios.
   * Se calculan campos derivados como TotalPrice en order\_details.
3. **Carga**: Los datos transformados se cargan en la base de datos MySQL ventas en las tablas products, customers, orders, y order\_details, reemplazando los datos existentes.

**Diagrama de Base de Datos**

El diagrama de la base de datos se puede generar usando una herramienta como MySQL Workbench:

* **Tablas**:
  + products: ProductID (PK), ProductName, Category, Price, Stock.
  + customers: CustomerID (PK), FirstName, LastName, Email, Phone, City, Country.
  + orders: OrderID (PK), CustomerID (FK), OrderDate, Status.
  + order\_details: OrderID (PK, FK), ProductID (PK, FK), Quantity, TotalPrice.
* **Relaciones**:
  + orders.CustomerID referencia customers.CustomerID.
  + order\_details.OrderID referencia orders.OrderID.
  + order\_details.ProductID referencia products.ProductID.

**DIAGRAMA**

Una captura de pantalla de una computadora

El contenido generado por IA puede ser incorrecto.

RESULTADOS

SELECT COUNT(\*) AS products\_count FROM products;

Interfaz de usuario gráfica, Texto, Aplicación

El contenido generado por IA puede ser incorrecto.