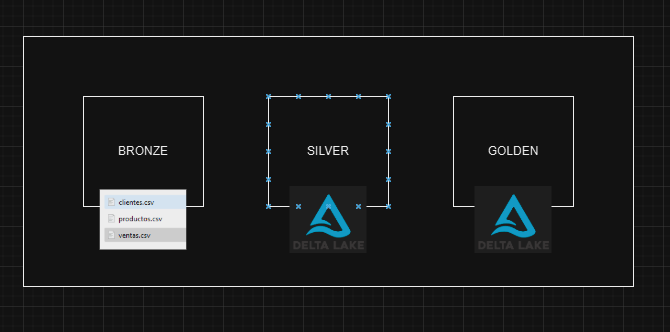
**Trabajo Final**

Implementar un datalake usando los csv proporcionados:



Capa Bronze: Guardar los archivos en el mismo formato de Origen.

Capa Silver: Crear tablas deltas a partir de los archivos guardados en la capa Bronze.

Capa Golden: Implementar los reportes finales en tablas Delta.

Detalle de las lógicas de negocio a implementar:

**1. Ventas Totales por Producto**

**Enunciado:**  
Realiza un reporte que indique el total de ventas realizadas para cada producto. El reporte debe mostrar:

* **Nombre del producto**
* **Categoría**
* **Total de cantidad vendida**  
  Para ello, deberás realizar un **join** entre los datasets ventas.csv y productos.csv utilizando la columna id\_producto. Luego, debes agrupar los datos por el nombre y la categoría del producto y sumar la cantidad vendida.

**2. Productos y sus Precios con Ventas Realizadas**

**Enunciado:**  
Crea un reporte que muestre los productos junto con su precio y la cantidad vendida, así como el total de ingresos generados por cada producto. El reporte debe incluir:

* **Nombre del producto**
* **Precio del producto**
* **Cantidad vendida**
* **Total de ingresos generados (cantidad vendida \* precio)**  
  Para esto, realiza un **join** entre los datasets productos.csv y ventas.csv utilizando la columna id\_producto. Posteriormente, calcula el total de ingresos generados por cada producto.

**3. Clientes con la Mayor Cantidad de Compras**

**Enunciado:**  
Elabora un reporte que muestre a los clientes que han realizado el mayor número de compras. El reporte debe incluir:

* **Nombre del cliente**
* **Cantidad de compras realizadas**  
  Para ello, realiza un **join** entre los datasets clientes.csv y ventas.csv utilizando la columna id\_cliente. Luego, agrupa por cliente y cuenta la cantidad de compras realizadas.

**4. Productos con Baja Venta y Stock Disponible**

**Enunciado:**  
Genera un reporte con los productos que tienen una baja venta, y que además tienen mucho stock disponible sin vender. El reporte debe incluir:

* **Nombre del producto**
* **Categoría**
* **Cantidad vendida**
* **Cantidad de stock disponible**  
  Para esto, debes realizar un **join** entre los datasets productos.csv y ventas.csv utilizando id\_producto. Luego, calcula las ventas totales por producto y compáralas con el stock disponible. Filtra aquellos productos que no han vendido muchas unidades y tienen alto stock restante.

Para la implementación del trabajo final pueden utilizar PySpark o Scala, utilizar databricks community.

**Fecha de entrega 9 de enero del 2025 – Hora límite 23:59 pm**