

***Nota: Nos equivocamos y le pusimos a todo lab4 pero en realidad debimos haber puesto lab3.**

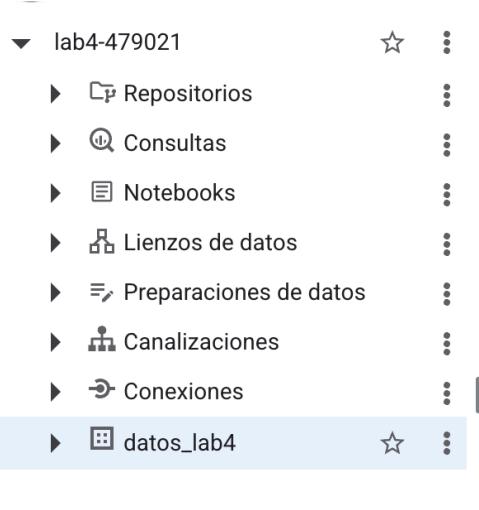
Creación del Bucket en Cloud Storage

The screenshot shows the Google Cloud Storage interface. A modal window titled 'Fija los buckets de uso frecuente' (Pin frequent-use buckets) is open, providing instructions on how to pin frequently used buckets for easier access. The main pane displays the 'lab4bucketbi' bucket details, including 'Protección' (Protection) set to 'No público' (Public), and a list of objects: OrderLines.csv, Orders.csv, PackageType.csv, and StockItems.csv. A confirmation message at the bottom states 'Se borraron 5 buckets' (5 buckets deleted).

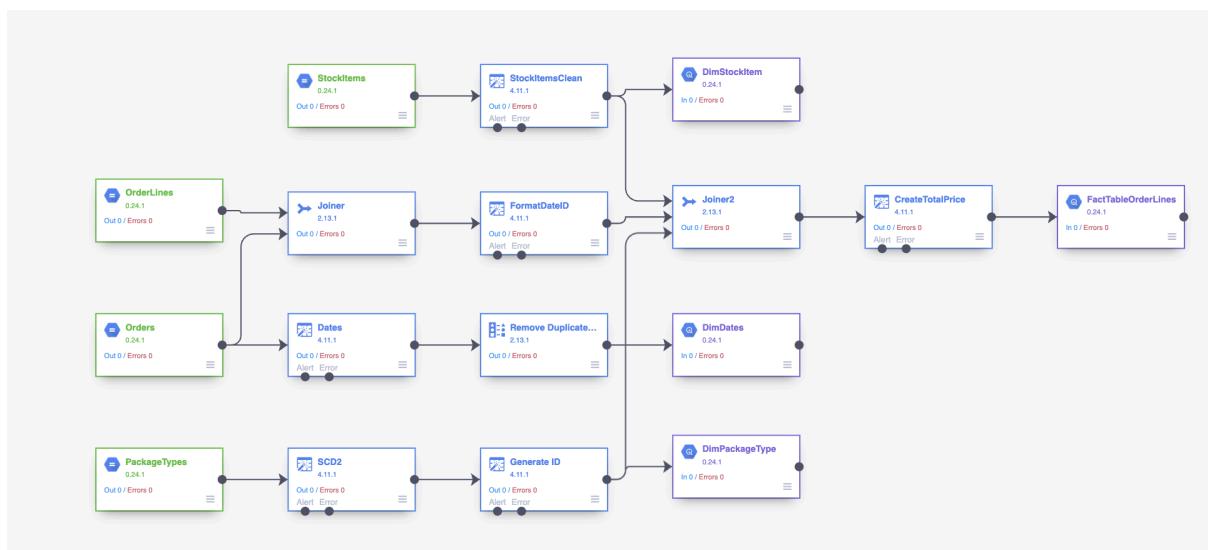
Creación del Data Fusion

The screenshot shows the Google Cloud Data Fusion interface. A search bar at the top contains the text 'data fusion'. Below it, a table lists existing instances: 'Instancias' (Instances) with one entry, 'instanciaLab4', which is selected (indicated by a green checkmark). The table columns include 'Nombre de la instancia' (Instance name), 'Acción' (Action), 'Edición' (Edition), 'Región' (Region), 'Zona' (Zone), 'Versión' (Version), 'Notificaciones' (Notifications), and 'Encriptación' (Encryption). The 'instanciaLab4' row shows 'Ver instancia' (View instance) under 'Acción'.

Creación del Big Query



Pasos de creación del proceso del ETL



El diagrama presentado evidencia la arquitectura completa del proceso ETL, el cual inicia con la extracción de las fuentes transaccionales (Orders, OrderLines, StockItems, PackageTypes y la tabla de fechas generada). A partir de estas fuentes, el flujo avanza hacia diferentes transformaciones intermedias que preparan los datos para su posterior carga en las dimensiones y la tabla de hechos del modelo multidimensional. En la parte superior del pipeline se observa el proceso relacionado con StockItems, el cual incluye el nodo StockItemsClean, encargado de normalizar y estandarizar atributos como nombres, precios, tasas de impuesto y claves de negocio. Posteriormente, esta información se dirige

hacia la dimensión DimStockItem, donde se almacena cada producto con sus atributos descriptivos.

En paralelo, los datos provenientes de OrderLines se integran con los de Orders mediante el nodo Joiner, permitiendo recuperar la fecha de la orden y otros atributos operacionales. A continuación, el nodo FormatDateID genera un identificador de fecha estandarizado (DateID). Este identificador permite mantener la granularidad diaria en la dimensión Date. Tras eliminar duplicados en el flujo de fechas y generar la dimensión DimDates, el pipeline garantiza que cada transacción se asigne correctamente a su día correspondiente.

Una de las transformaciones más relevantes del proceso es la unión final realizada en el nodo Joiner2, donde convergen las líneas de orden, la información del producto (StockItem) y la información del empaque (PackageType). Este último abastece la nueva dimensión DimPackageType, la cual incorpora un manejo de historia tipo 2 (SCD2). Este manejo implica generar versiones históricas de los atributos del empaque. De esta forma, si un PackageType cambia alguno de sus valores, se registra una nueva fila con un surrogate key independiente, y se controla la vigencia con los atributos:

- StartDate
- EndDate
- IsCurrent

Esto asegura que las medidas históricas se mantengan consistentes con el contexto temporal en el que fueron registradas.

Además, se añadió una nueva medida calculada llamada TotalPrice, definida a partir de la fórmula:

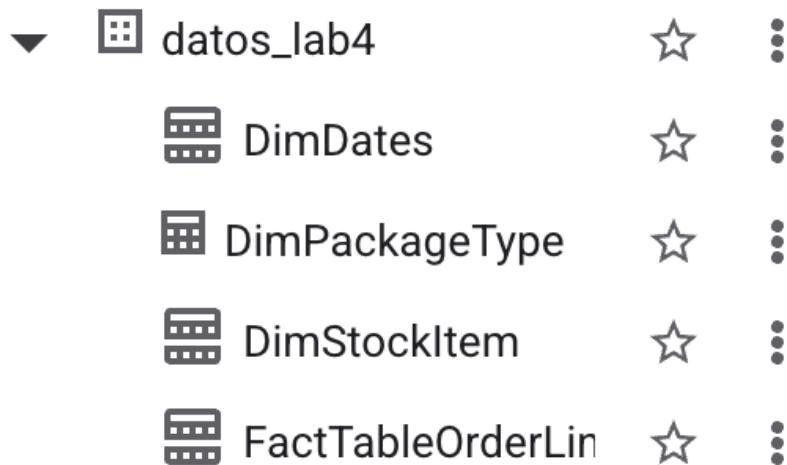
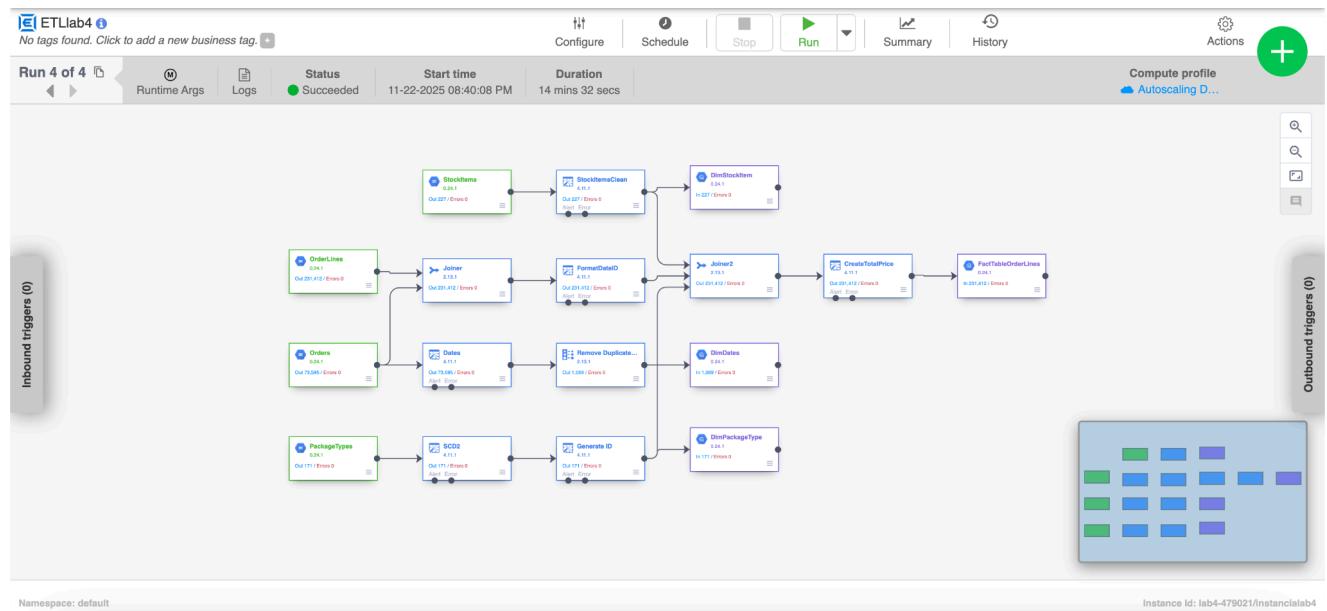
$$\text{TotalPrice} = \text{Quantity} \times \text{UnitPrice}/\text{PriceLast}$$

Esta transformación se realiza en el nodo CreateTotalPrice. La medida permite analizar los ingresos brutos por línea de pedido, lo cual es uno de los requerimientos de negocio planteados en el caso de estudio.

Finalmente, toda la información converge en la tabla de hechos FactTableOrderLines, que contiene:

- Las medidas: Quantity, PickedQuantity, UnitPrice, TaxRate y la nueva TotalPrice.
- Las claves foráneas hacia las dimensiones Date, StockItem y PackageType.
- Los identificadores degenerados: OrderID y OrderLineID, que se mantienen en la tabla de hechos al no contar con atributos propios que ameriten una dimensión independiente.

Evidencia de ejecución correcta del proceso ETL



Como se puede observar, el ETL se ejecuto correctamente y se crearon las tablas en BigQuery (en el conjunto de datos “datos_lab4”).