

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

Creación del Bucket en Cloud Storage

lab4bucketbi

Fija los buckets de uso frecuente

Ahora puedes fijar los buckets a los que accedes con frecuencia en la parte superior de la página de la lista de buckets, lo que te permitirá encontrarlos rápidamente en el futuro.

lab4bucketbi

Nombre

Acceso público

Protección

No público

Borrar de forma no definitiva

Protección

Ciclo de vida

Observabilidad

Nuevo

Informes de inventario

Operaciones

Depósitos

lab4bucketbi

Crear carpeta

Subir

Transferir los datos

Otros servicios

Filtrar solo por prefijo de nombre

Filtro

Filtrar objetos y carpetas

Mostrar

Solo objetos activos

Nombre	Tamaño	Tipo	Fecha de creación	Clase de almacén
OrderLines.csv	15.8 MB	text/csv	22 nov 2025, 4:04:59 p.m.	Standard
Orders.csv	8.2 MB	text/csv	22 nov 2025, 4:04:58 p.m.	Standard
PackageType.csv	7.4 KB	text/csv	22 nov 2025, 4:04:52 p.m.	Standard
StockItems.csv	69.2 KB	text/csv	22 nov 2025, 4:04:52 p.m.	Standard

Se borraron 5 buckets

Creación del Data Fusion

data fusion

Buscar

Instancias

Crear instancia

Actualizar

Borrar

Actualizar

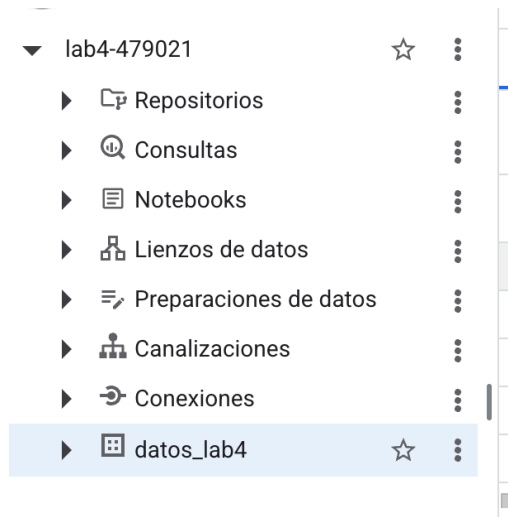
¿Cómo fue tu experiencia?

Más información

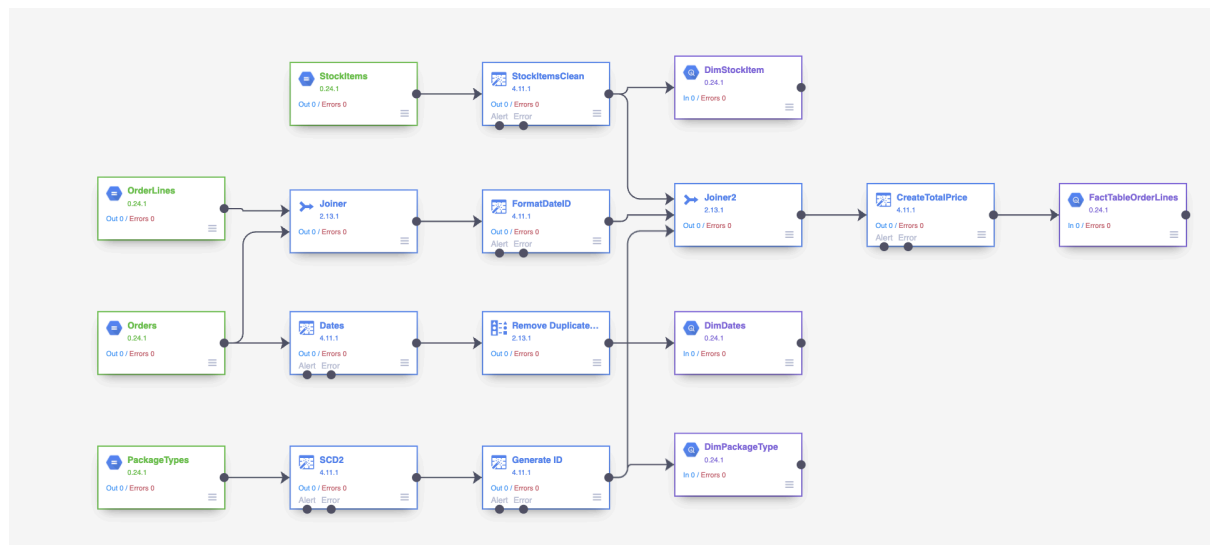
Selecciona la instancia de Cloud Data Fusion que deseas ver.

Nombre de la instancia	Acción	Edición	Región	Zona	Versión	Notificaciones	Encriptación
instanciaLab4	Ver instancia	Basic	us-west1	--	6.11.1 (latest version)		Administrada por Google

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:

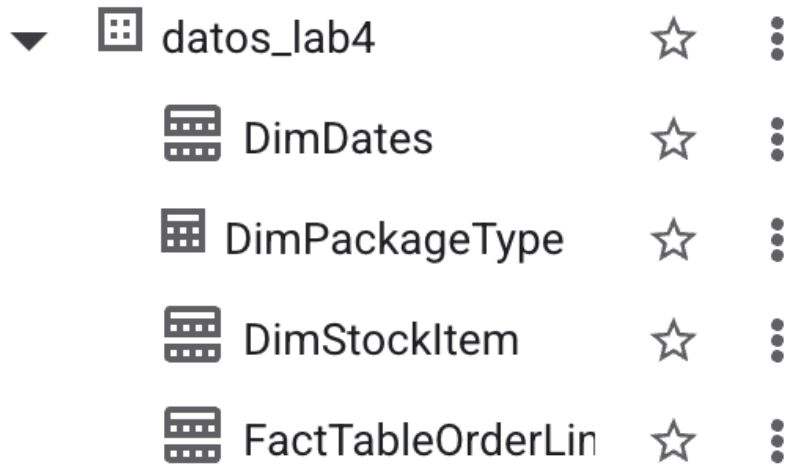
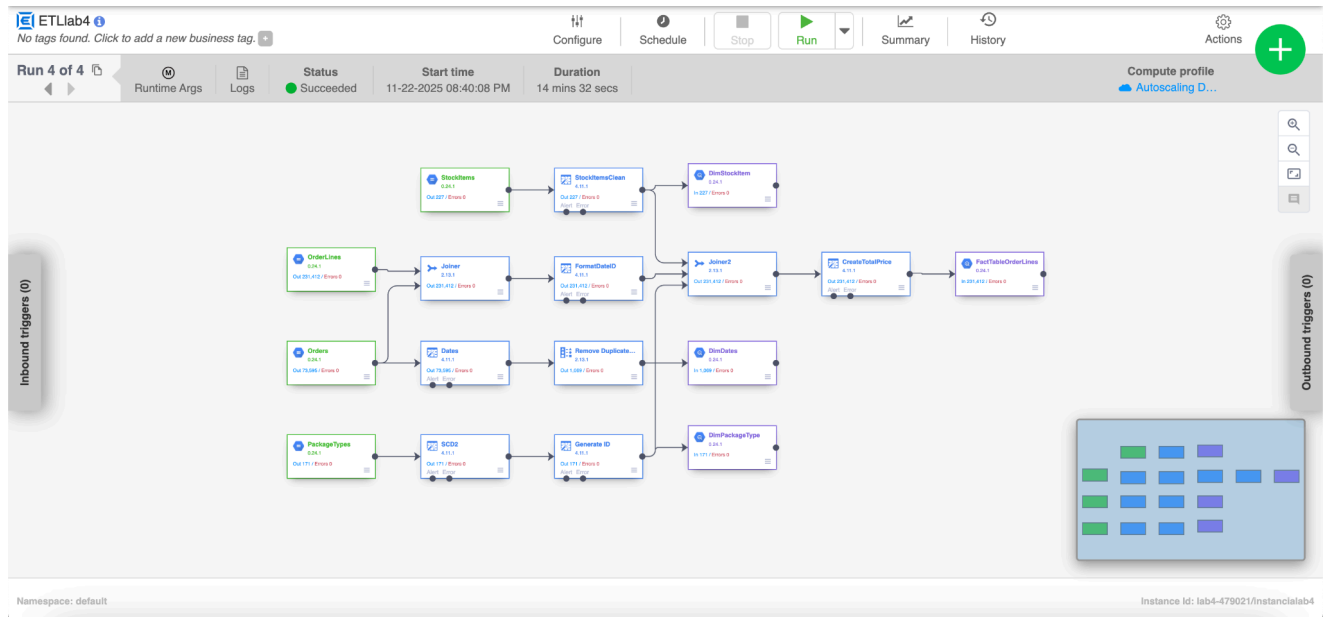
TotalPrice = Quantity x UnitPrice/PriceLast

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”).