

**Evidencia de aprendizaje 3. Proceso de transformación de datos y carga en el data mart  
final**

**Bases de Datos II**

**Ricardo García Duque  
Edison Fabián Molina Herrera  
Yamith Alejandro Castaño Pérez**

Profesor:

**Victor Hugo Mercado Ramos**

Grupo.

**PREICA2401B010094**

**Institución Universitaria Digital de Antioquia**

**2024.**

## INTRODUCCIÓN

En el mundo de la inteligencia de negocios, estamos continuamente en la búsqueda y construcción de mecanismos que brinden de manera más simple, rápida, eficiente y segura datos estructurados, información real, verídica y cuyo contexto se aplique a múltiples realidades actuales, es por ello que hoy abordamos la temática de la construcción de modelos dimensionales y su organización en esquemas como el modelo estrella. En este tema observaremos cómo se han dimensionado los aspectos relevantes de la base de datos suministrada, de qué manera se han discriminado en campos y como se han organizado para generar tablas dimensionales las cuales convergen en una tabla de hechos acorde a la realidad del negocio.

En desarrollo de este documento se verán los procesos y pasos que conllevaron a la creación final del DataMart, pasando desde el proceso inicial que fue el planteamiento del modelo estrella, clasificar información relevante y necesaria para posteriormente crear una base de datos Staging, hacer el proceso de transformación de la información y finalmente hacer la transferencia y carga de datos en el DataMart final.

## OBJETIVOS

- Aplicar los conocimientos adquiridos en el desarrollo del curso asociados al diseño de un Data mart bajo el modelo estrella.
- Realizar el proceso de transformación y carga de datos en el datamart final.
- desarrollar habilidades y adquirir conocimientos que serán la base para nuestro desarrollo profesional.

## PLANTEAMIENTO DEL PROBLEMA

Tenemos una base de datos que describe el proceso de negocio de una empresa que comercializa productos del agro, herramientas, plantas, árboles, utensilios para maquinaria agrícola, etc.

Deseamos crear un modelo de datos que nos permita visualizar el total de la información de manera rápida y ágil dando una perspectiva eficaz desde el total de ventas hasta el stock del inventario de productos y así poder tener un panorama claro de todas las actividades comerciales de la empresa, lo que permitiría establecer lineamientos claros sobre las decisiones estratégicas como los productos más vendidos, los menos comercializados, la reducción de stock de los productos menos vendidos, la afectación positiva o negativa de las temporadas de cosecha en los tipos de árboles y plantas que se venden, la inclinación de los clientes hacia ciertos tipos de marcas, etc.

Esta información sería primordial para generar un impacto positivo en el rumbo de los procesos comerciales de la empresa, para ello debemos enfocar el análisis en 3 aspectos esenciales:

- Los productos más vendidos.
- La categoría que tiene más productos, tanto en ventas como en stock de inventario.
- El periodo de tiempo con más ventas y tipo de ventas en ese lapso de tiempo.

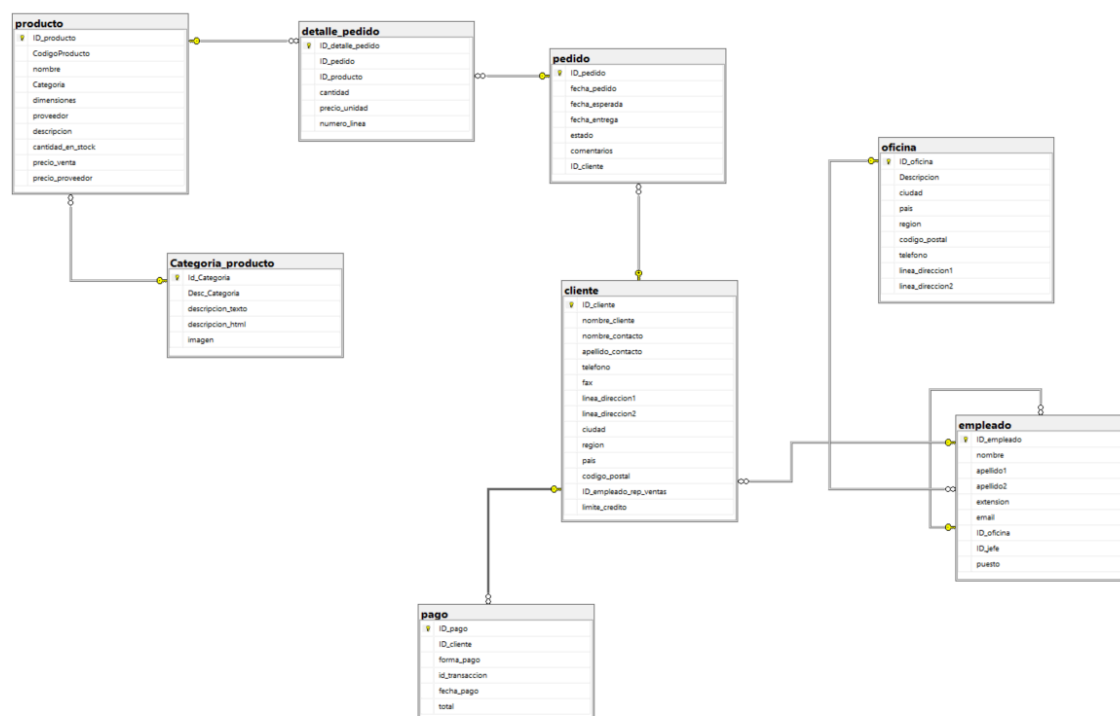
En esencia, el problema se constituye en la ausencia de un modelo que proporcione información estructurada y que brinde conocimiento contextualizado, real y eficaz que posibilite la toma de decisiones y el mejoramiento de las actividades comerciales de la empresa.

## ANÁLISIS DEL PROBLEMA

Jardinería “La parcela” cuenta con una gran cantidad de datos históricos de los últimos 8 años de su operación, los cuales, aunque inicialmente estaban dispersos en archivos planos como excel, fueron trasladados a una base de datos bajo el modelo relacional utilizando mongoDB, pero aún resulta complejo poder relacionar y obtener los datos necesarios para identificar resultados o falencias en los procesos de la empresa.

La información se encuentra en diferentes tablas y se presenta en la siguiente imagen.

Imágen del modelo relacional.



Jardinería “La parcela” como ya fue planteado requiere la construcción de una solución que le permita realizar su propia inteligencia de negocios y con ello, acceder de una manera confiable a la toma de decisiones con base en información de valor que hoy no es posible obtener de manera rápida y eficiente.

## PROPUESTA DE LA SOLUCIÓN

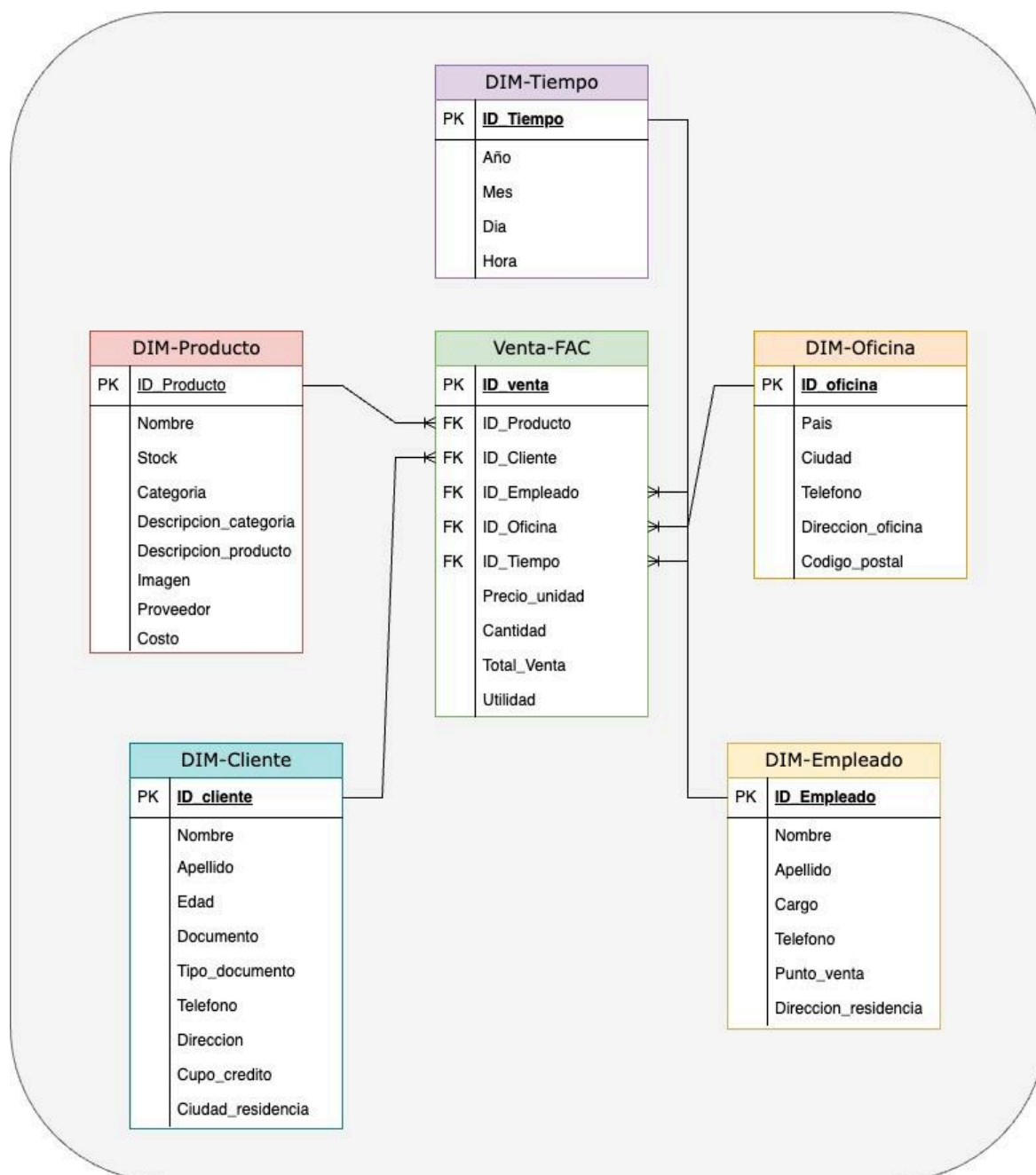
- **Descripción del modelo.**

Es un modelo que permite correlacionar datos que existen en diferentes fuentes para hacerlos accesibles de manera más ágil. Tiene como fundamento la existencia de una tabla de hechos que almacena datos cuantitativos y está rodeada de dimensiones o tablas con datos o atributos descriptivos.

Para la propuesta del modelo se definieron dimensiones que satisfacen las necesidades en cuanto a la información necesaria para poder responder ante las 3 necesidades del problema permitiendo identificar el producto más vendido, la categoría con más productos y el año con más ventas utilizando la base de datos Jardinería como fuente de datos.

De lo anterior surge como resultado la tabla de hechos, eje central de la solución que permitirá la consulta de la información de una manera ágil y centralizada para poder monitorear y consultar datos específicos así como poder predecir comportamientos del negocio que permitan la toma de futuras decisiones.

- **Diseño.**



- **Lista de dimensiones**

- **Dimensión tiempo**

DIM-Tiempo	
PK	<u>ID_Tiempo (int)</u>
	Año (time)
	Mes (time)
	Dia (time)
	Hora (time)

- **Dimensión producto**

DIM-Producto	
PK	<u>ID_Producto (int)</u>
	Nombre (varchar)
	Stock (int)
	Categoria (varchar)
	Descripcion_categoria (texto)
	Descripcion_producto (texto)
	Imagen (varchar)
	Proveedor (varchar)
	Costo (varchar)

- **Dimensión cliente**

DIM-Cliente	
PK	<u>ID_cliente (int)</u>
	Nombre (varchar)
	Apellido (varchar)
	Edad (varchar)
	Documento (int)
	Tipo_documento (varchar)
	Telefono (int)
	Direccion (varchar)
	Cupo_credito (int)
	Ciudad_residencia (varchar)



■ Dimensión oficina

DIM-Oficina	
PK	<u>ID_oficina (int)</u>
	Pais (varchar)
	Ciudad (varchar)
	Telefono (int)
	Direccion_oficina (varchar)
	Codigo_postal (int)

■ Dimensión empleado

DIM-Empleado	
PK	<u>ID_Empleado (int)</u>
	Nombre (varchar)
	Apellido (varchar)
	Cargo (varchar)
	Telefono (int)
	Punto_venta (varchar)
	Direccion_residencia (varchar)

- **Tabla de hechos**

Venta-FAC	
PK	<u>ID_venta (int)</u>
FK	ID_Producto (int)
FK	ID_Cliente (int)
FK	ID_Tiempo (int)
FK	ID_Oficina (int)
FK	ID_Empleado (int)
	Precio_unidad (float)
	Cantidad (int)
	Total_Venta (float)
	Utilidad (float)

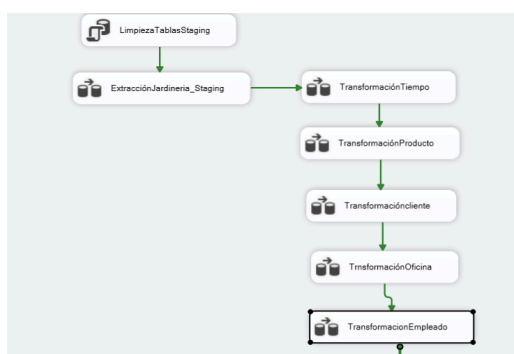
La tabla de hechos como centro del modelo estrella, contiene las llaves foráneas de todas las dimensiones permitiendo la relación entre los datos que las habitan tales como producto, cliente, tiempo, oficina, empleado y permite consolidar datos como precio\_unidad, cantidad, total\_venta y utilidad.

Además contiene medidas de tipo numérico, facilitando la sumarización entre ellas y entre los atributos de las dimensiones. Con lo cual se consigue obtener información ágil y contextual del negocio, que en suma es el objetivo central del modelo dimensional.

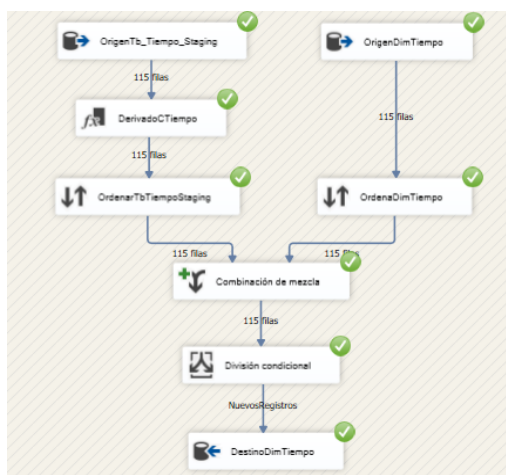
## Creación de las dimensiones

A continuación se muestran los detalles de la transformación realizada luego de haber extraído los datos en la base de datos Staging Jardinería.

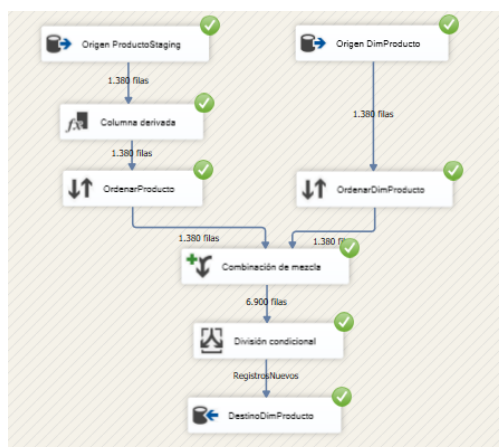
### Flujo de control



### Transformación hacia la Dimensión de Tiempo



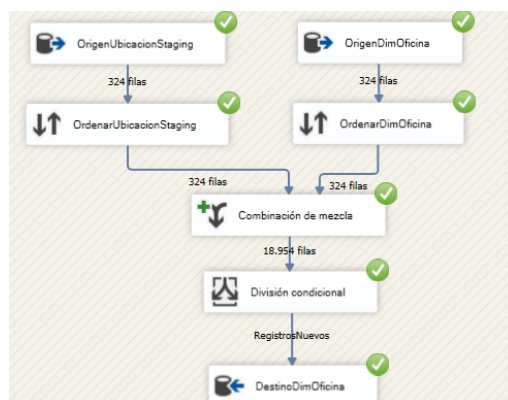
### Transformación hacia la Dimensión de Producto



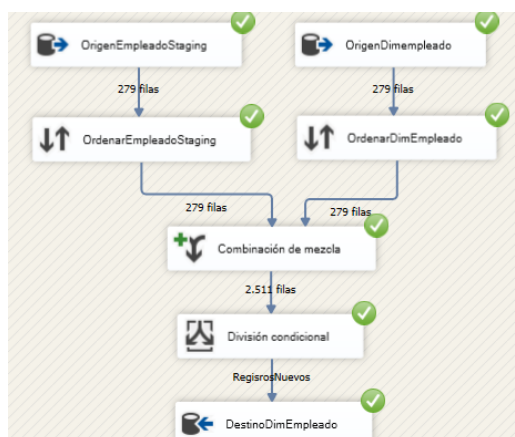
### Transformación hacia la Dimensión de Cliente



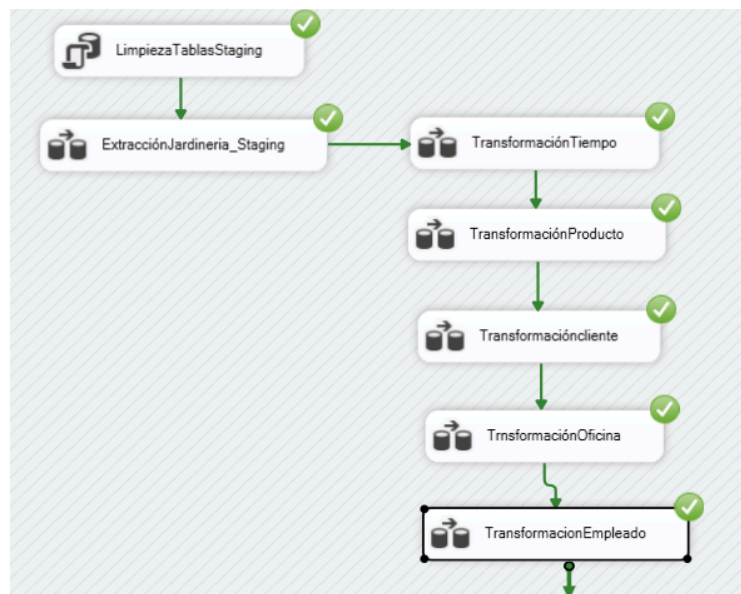
### Transformación hacia la Dimensión de Oficina



### Transformación hacia la Dimensión de Empleado



### Resultado final



### DataMart Final

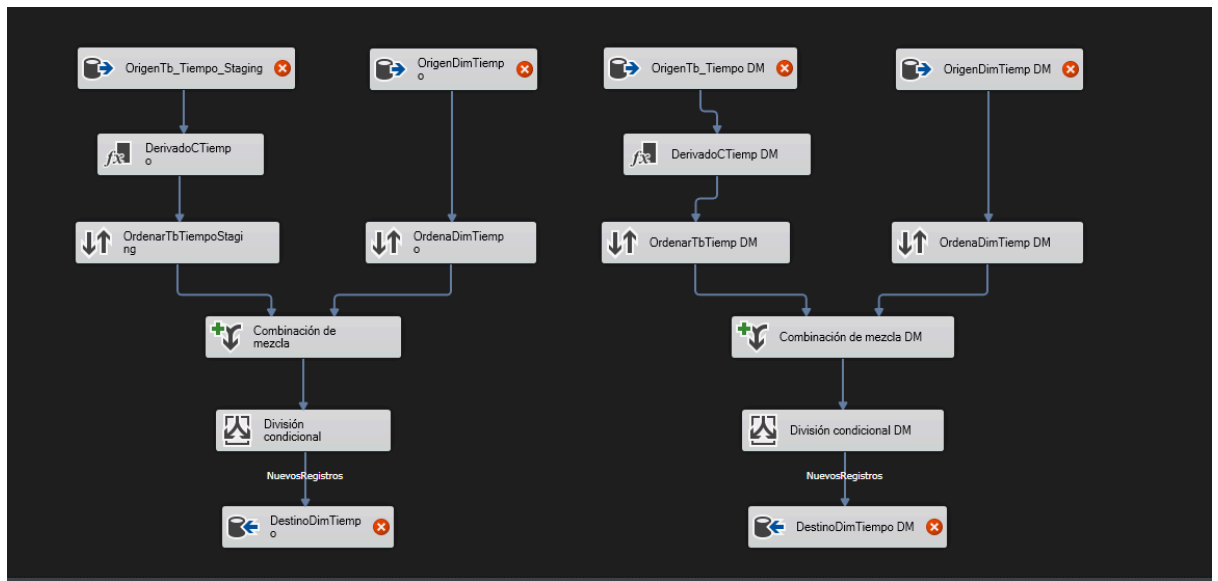
En el proceso para la carga final al DataMart se tuvieron 3 aspectos importantes:

1. Extracción de los datos: en este punto se recopiló información suministrada por la base de datos jardinería utilizando las tablas que fueron definidas en el modelo estrella.
2. Transformación : Los datos extraídos se sometieron a procesos de limpieza, agregación y transformación permitiendo que sean de uso adecuado para proceder con el análisis de la información, en este punto intervienen procesos de

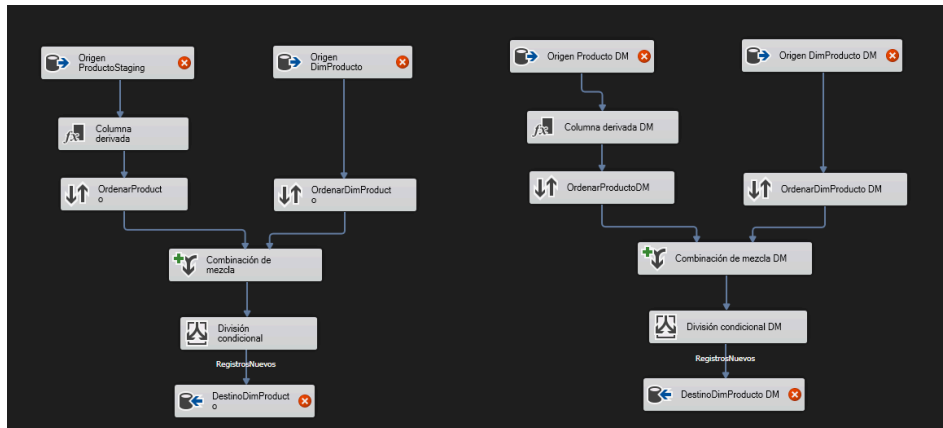
### Esquema general.



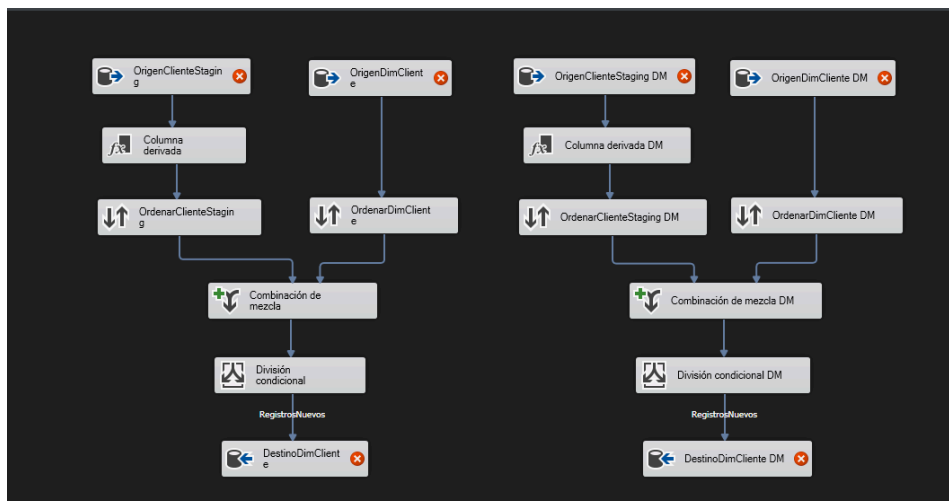
Se realiza la transformación , derivación, ordenamiento, combinación y división datos justo a lo trabajado para el proceso Staging.



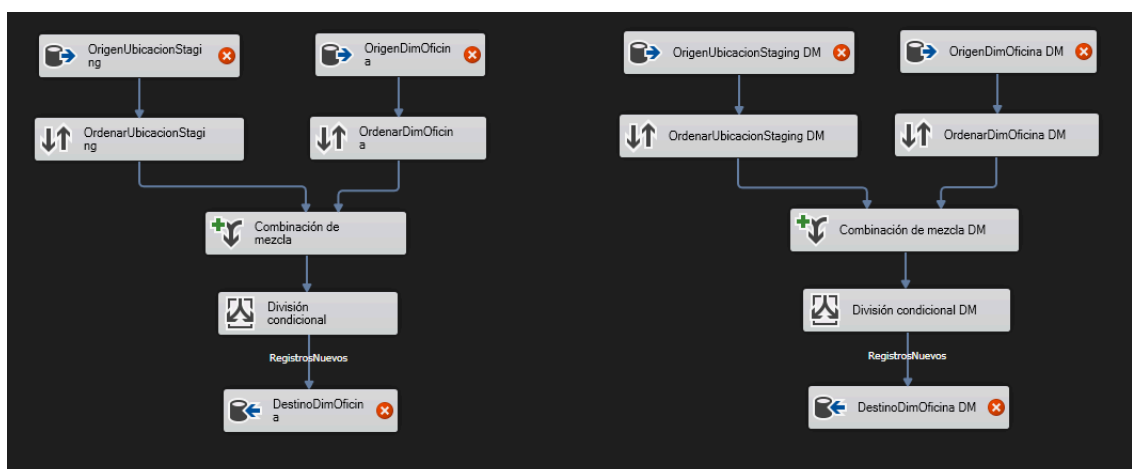
Producto:



Cliente:

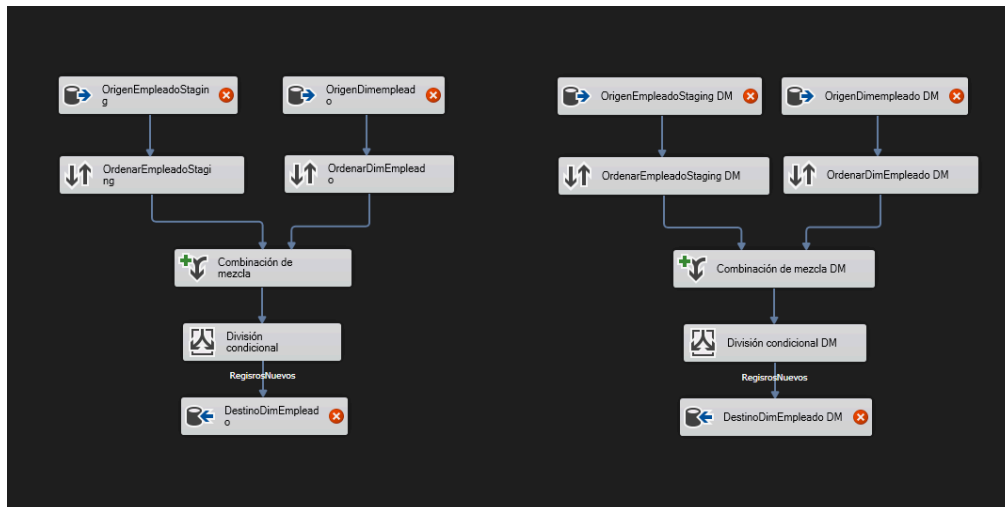


Oficina:

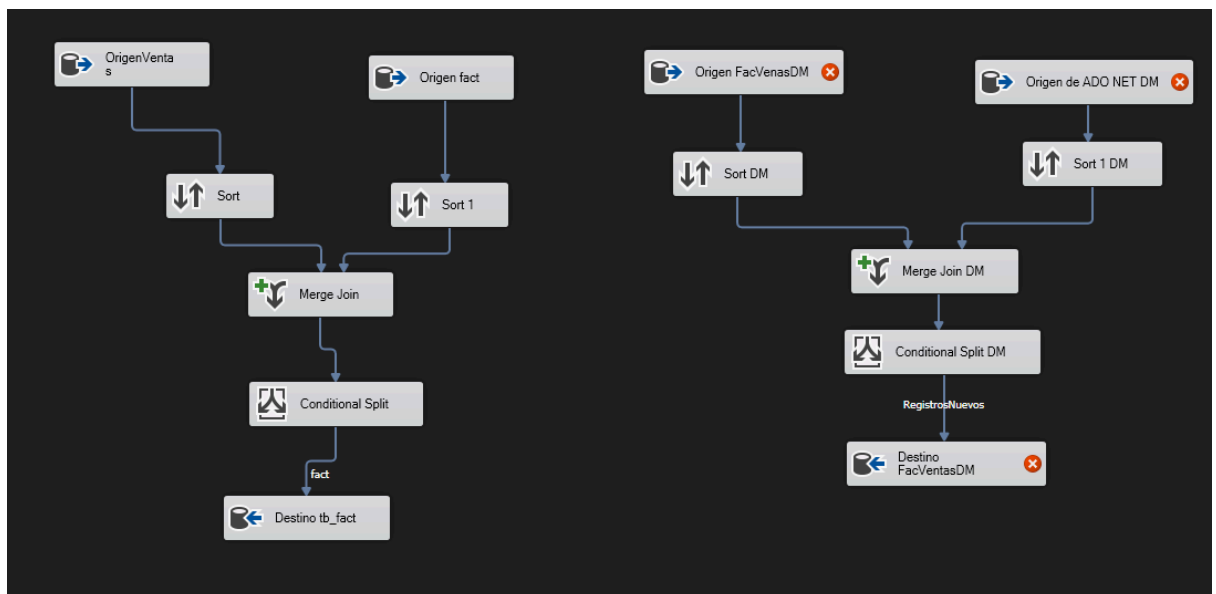




Empleado:



Hechos:



## CONCLUSIONES:

- Resaltar la importancia de la creación y justificación del modelo estrella como una herramienta actual, funcional y de gran impacto en la capacidad para la toma de decisiones gerenciales, estrategias comerciales y para el crecimiento y alcance de la compañía.
- Hemos diseñado y construido un modelo estrella para un *data mart* que permita analizar y responder a tres categorías específicas: identificar el producto más vendido, la categoría con más productos y el año con más ventas, utilizando la base de datos Jardinería como fuente de datos.
- Luego de realizar el proceso de transformación mediante técnicas de ETL y sentencias SQL se observó que los datos y las tablas presentan una mejor apreciación y generan información confiable.
- La carga de datos al Data Mart final se logró realizar gracias a todo lo visto en clase, a la información en los módulos y consultas en internet, lo que nos deja una experiencia importante, conocimientos base que serán muy importante en el desarrollo de nuestra carrera.
- Al finalizar este ejercicio, hemos adquirido experiencia práctica en el diseño y desarrollo de procesos ETL utilizando herramientas como SQL Server y Microsoft Visual Studio.

## Bibliografía

- Alfaro, F. R. (2021, 17 junio). *El modelo de estrella. El pilar fundamental del Business Intelligence - Data Management*. Data Management. <https://datamanagement.es/2019/06/27/business-intelligence-modelo-estrella/>
- IBM documentation. (s. f.). <https://www.ibm.com/docs/es/ida/9.1.2?topic=schemas-star>
- Manuel. (2023, 31 julio). *Qué es modelo de datos en estrella*. The Data Schools. <https://thedata.schools.com/que-es/modelo-de-datos-estrella/>
- Peter-Myers. (2023, 23 marzo). *Descripción de un esquema de estrella e importancia para Power BI - Power BI*. Microsoft Learn. <https://learn.microsoft.com/es-es/power-bi/guidance/star-schema>
- *Data warehouse technology options - Data Warehousing on AWS*. (s. f.). <https://docs.aws.amazon.com/whitepapers/latest/data-warehousing-on-aws/data-warehouse-technology-options.html>
- *¿Qué es un data mart? | Oracle*. (s. f.). <https://www.oracle.com/co/autonomous-database/what-is-data-mart/>