

**S30 - Evidencia de aprendizaje 3. Proceso de transformación de datos  
y carga en el data mart final**

**Juan David Maya Tobon**

**CC 1.035.831.285**

**Grupo 94**

**Docente**

**Victor Hugo Mercado Ramos**

**Bases de Datos II**

**IU Digital de Antioquia**

**Ingeniería de Software y Datos**

**25 de mayo de 2024**

## **Introducción**

En el ámbito de la gestión de bases de datos, el modelo estrella emerge como un pilar fundamental en el diseño multidimensional, especialmente valioso para estructurar datos en entornos empresariales complejos, este enfoque adquiere una importancia notable en sistemas de información que demandan un análisis profundo de datos, como las transacciones comerciales de empresas, donde la organización eficiente y el acceso ágil a la información se vuelven críticos, en este estudio, profundizaremos en la comprensión y aplicación del modelo estrella dentro del contexto de una base de datos enfocada en el sector de la jardinería, exploraremos sus elementos clave y evaluaremos su capacidad para mejorar la gestión de datos empresariales en este sector específico, destacando su potencial para impulsar decisiones estratégicas informadas y optimizar procesos de análisis de información.

Además, abordaremos los procesos de carga y transformación de datos, esenciales para garantizar que la información almacenada en el modelo estrella sea precisa y esté debidamente estructurada, estos procesos incluyen la extracción de datos desde diversas fuentes, su transformación para cumplir con los estándares y requisitos del modelo, y su carga en la base de datos final, este enfoque garantiza que los datos sean consistentes y confiables, lo cual es crucial para realizar análisis precisos y tomar decisiones bien fundamentadas; analizaremos cómo estas prácticas de ETL (Extracción, Transformación y Carga) se implementan específicamente en el sector de la jardinería, contribuyendo a una mejor gestión y utilización de la información.

## Objetivos

Diseñar e implementar la estructura de un modelo estrella en la base de datos de la empresa de jardinería, optimizando la gestión y análisis de datos organizacionales.

Establecer relaciones claras entre la tabla de hechos de ventas y las dimensiones pertinentes, como clientes, productos, tiempo y pagos, garantizando la integridad referencial y la consistencia de la información.

Mejorar la eficiencia en la consulta y extracción de datos mediante la implementación de índices y optimización de consultas, permitiendo un acceso rápido y preciso a la información requerida para análisis y reportes.

Establecer una estructura organizada y coherente en la base de datos, facilitando el análisis y la gestión de la información relacionada con ventas, clientes, productos y transacciones.

Implementar y evaluar procesos de transformación y carga de datos (ETL) en un modelo estrella para una base de datos del sector de la jardinería, y así de esta manera asegurar que los datos transformados sean precisos y consistentes, optimizando así la gestión y análisis de la información para mejorar la toma de decisiones estratégicas.

### **Planteamiento del problema**

La necesidad imperante de optimizar la gestión de información en la base de datos de la empresa de jardinería, surge tras la ausencia de una estructura organizada, relaciones claras entre los datos y la presencia de inconsistencias; aspectos que dificultan el análisis de ventas, seguimiento de clientes y toma de decisiones estratégicas. Dichos factores, tras ser analizados plantean el objetivo de desarrollar e implementar un modelo estrella que permita gestionar de manera eficiente los datos de ventas, clientes, productos y transacciones; con el fin de mejorar la precisión y actualización de la información, facilitando así, la toma de decisiones fundamentadas en datos confiables y oportunos.

## **Análisis del problema**

La implementación del modelo estrella surge de la necesidad de mejorar la gestión de la información en la base de datos de la empresa de jardinería, donde actualmente, la estructura de la base de datos presenta ciertas deficiencias en términos de organización y relaciones de datos; aspectos que dificultan el análisis integral de las ventas, el seguimiento detallado de los clientes y la toma de decisiones estratégicas; lo que impacta directamente la capacidad de la empresa para obtener información precisa y oportuna sobre sus operaciones comerciales, afectando a su vez, la eficiencia en la toma de decisiones y el rendimiento general del negocio.

Al adoptar el modelo estrella, se busca establecer una estructura más robusta y coherente en la base de datos, ya que este modelo se caracteriza por tener una tabla de hechos central que registra las transacciones comerciales clave, como las ventas, y dimensiones relacionadas que contienen atributos descriptivos detallados, como información de clientes, productos y tiempos; lo que permite una organización más eficiente de los datos y una relación clara entre ellos, facilitando así la extracción de información relevante para análisis y reportes.

Con la implementación del modelo estrella, se espera mejorar significativamente la capacidad de la empresa para analizar y comprender sus datos comerciales, lo que incluye la capacidad de identificar tendencias de ventas, segmentar clientes de manera más precisa, evaluar el rendimiento de productos específicos y realizar proyecciones financieras más precisas; en última instancia, el objetivo es mejorar la toma de decisiones estratégicas y optimizar los procesos internos, aumentando la competitividad y el éxito general de la empresa.

## **Propuesta de la solución**

### **Descripción del modelo estrella propuesto**

El modelo estrella propuesto para la base de datos JardineriaModeloEstrella, es una estructura organizativa avanzada, que sigue una metodología de diseño multidimensional; la cual es crucial para gestionar eficientemente grandes volúmenes de datos, especialmente en el contexto de las complejas transacciones de ventas de una empresa de jardinería.

### ***Dimensiones Propuestas***

**Dimensión Tiempo:** Esta dimensión aborda el aspecto temporal de las transacciones de ventas, proporcionando campos como fecha\_pedido, anio, mes, trimestre y dia\_semana, estos atributos permiten análisis temporales detallados, como tendencias de ventas estacionales y comportamiento de compra a lo largo del tiempo.

**Dimensión Cliente:** Aquí se recopila información detallada sobre los clientes, incluyendo ID\_cliente, nombre\_cliente, datos de contacto, dirección y límite de crédito, lo que facilita la segmentación de clientes, seguimiento de relaciones comerciales y evaluación del rendimiento de ventas por cliente.

**Dimensión Producto:** Esta dimensión describe los productos vendidos, proporcionando datos como ID\_producto,CodigoProducto, nombre, categoría, dimensiones, proveedor, precios y stock, estos detalles son esenciales para analizar la popularidad de productos, tendencias de inventario y relación con proveedores.

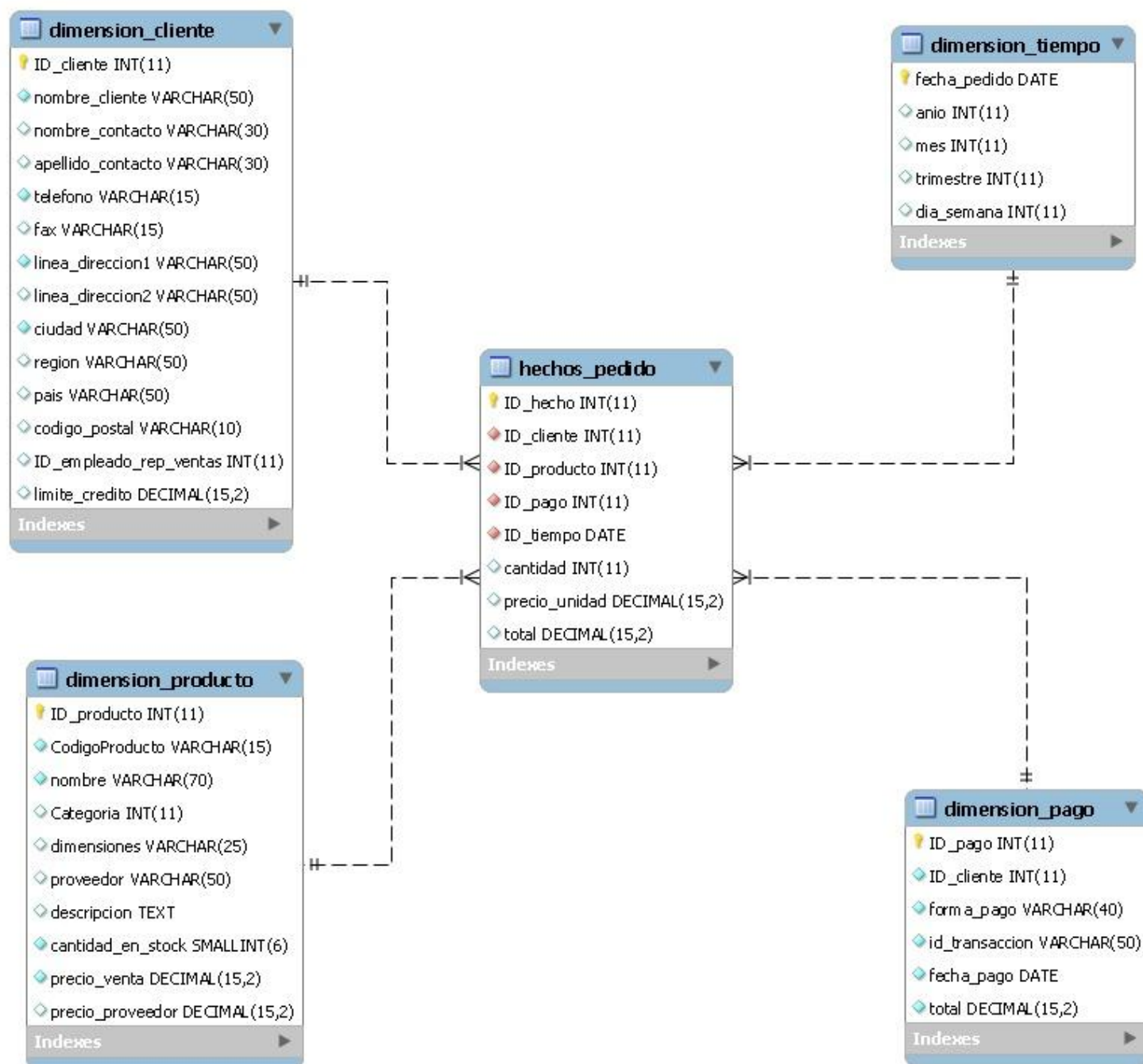
**Dimensión Pago:** Aquí se registra información financiera crucial asociada a las transacciones, incluyendo ID\_pago, ID\_cliente, forma de pago, ID de transacción, fecha y

total pagado, esta dimensión es fundamental para análisis financieros y seguimiento de pagos por cliente y período de tiempo.

**La tabla de hechos (hechos\_pedido):** Actúa como el núcleo central que integra las dimensiones mencionadas anteriormente, debido a que registra detalles clave de cada transacción de pedido, como ID\_cliente, ID\_producto, ID\_pago, ID\_tiempo, cantidad, precio\_unitario y total; la implementación de restricciones de integridad referencial garantiza la consistencia y coherencia de los datos entre las dimensiones y la tabla de hechos.

Este diseño del modelo estrella proporciona una estructura sólida y eficiente para realizar consultas complejas y análisis multidimensionales; lo que permite a la empresa de jardinería, comprender mejor sus operaciones comerciales, identificar tendencias de ventas, analizar el comportamiento del cliente y tomar decisiones estratégicas informadas basadas en datos precisos y oportunos.

## Diseño (Imagen) del modelo estrella





## Lista de dimensiones propuestas

### *Dimensión Tiempo*

**Campos:** fecha\_pedido, anio, mes, trimestre, dia\_semana.

**Motivo de uso:** Permite analizar las ventas y transacciones en función del tiempo, como identificar tendencias estacionales, patrones de compra mensuales o comparar el desempeño trimestral.

### *Dimensión Cliente*

**Campos:** ID\_cliente, nombre\_cliente, nombre\_contacto, apellido\_contacto, telefono, fax, linea\_direccion1, linea\_direccion2, ciudad, region, pais, codigo\_postal, ID\_empleado\_rep\_ventas, limite\_credito.

**Motivo de uso:** Facilita el análisis de ventas por cliente, permite segmentar clientes por ubicación geográfica, gestión de crédito, y asignar representantes de ventas.

### *Dimensión Producto*

**Campos:** ID\_producto, CodigoProducto, nombre, Categoria, dimensiones, proveedor, descripcion, cantidad\_en\_stock, precio\_venta, precio\_proveedor.

**Motivo de uso:** Permite analizar el desempeño de productos en ventas, gestionar el inventario, conocer proveedores y categorizar los productos para análisis detallados.

### *Dimensión Pago*

**Campos:** ID\_pago, ID\_cliente, forma\_pago, id\_transaccion, fecha\_pago, total.

**Motivo de uso:** Permite analizar los pagos asociados a las transacciones, identificar patrones de pago, formas de pago preferidas por los clientes y gestionar registros financieros.

## Dimensiones con las columnas y tipos de datos van a almacenar

### *Dimensión Tiempo*

**fecha\_pedido** DATE PRIMARY KEY COMMENT 'Fecha del pedido'

**anio** INT COMMENT 'Año del pedido'

**mes** INT COMMENT 'Mes del pedido'

**trimestre** INT COMMENT 'Trimestre del pedido'

**dia\_semana** INT COMMENT 'Día de la semana del pedido'

### *Dimensión Cliente*

**ID\_cliente** INT PRIMARY KEY

**nombre\_cliente** VARCHAR(50) NOT NULL COMMENT 'Nombre del cliente'

**nombre\_contacto** VARCHAR(30) COMMENT 'Nombre de contacto'

**apellido\_contacto** VARCHAR(30) COMMENT 'Apellido de contacto'

**telefono** VARCHAR(15) NOT NULL COMMENT 'Teléfono del cliente'

**fax** VARCHAR(15) COMMENT 'Fax del cliente'

**linea\_direccion1** VARCHAR(50) NOT NULL COMMENT 'Dirección línea 1'

**linea\_direccion2** VARCHAR(50) COMMENT 'Dirección línea 2'

**ciudad** VARCHAR(50) NOT NULL COMMENT 'Ciudad del cliente'

**region** VARCHAR(50) COMMENT 'Región del cliente'

**pais** VARCHAR(50) COMMENT 'País del cliente'

**codigo\_postal** VARCHAR(10) COMMENT 'Código postal del cliente'

**ID\_empleado\_rep\_ventas** INT COMMENT 'ID del empleado representante '

**limite\_credito** DECIMAL(15,2) COMMENT 'Límite de crédito del cliente'

### ***Dimensión Producto***

**ID\_producto** INT PRIMARY KEY,  
**CodigoProducto** VARCHAR(15) NOT NULL COMMENT 'Código del producto'  
**nombre** VARCHAR(70) NOT NULL COMMENT 'Nombre del producto'  
**Categoria** INT COMMENT 'ID de la categoría del producto'  
**dimensiones** VARCHAR(25) COMMENT 'Dimensiones del producto'  
**proveedor** VARCHAR(50) COMMENT 'Proveedor del producto'  
**descripcion** TEXT COMMENT 'Descripción del producto'  
**cantidad\_en\_stock** SMALLINT NOT NULL COMMENT 'Cantidad en stock'  
**precio\_venta** DECIMAL(15,2) NOT NULL COMMENT 'Precio de venta'  
**precio\_proveedor** DECIMAL(15,2) COMMENT 'Precio de proveedor del producto'

### ***Dimensión Pago***

**ID\_pago** INT PRIMARY KEY,  
**ID\_cliente** INT NOT NULL COMMENT 'ID del cliente asociado al pago'  
**forma\_pago** VARCHAR(40) NOT NULL COMMENT 'Forma de pago'  
**id\_transaccion** VARCHAR(50) NOT NULL COMMENT 'ID de transacción'  
**fecha\_pago** DATE NOT NULL COMMENT 'Fecha del pago'  
**total** DECIMAL(15,2) NOT NULL COMMENT 'Total del pago'

### **Tabla de hechos, con sus campos y tipos de datos**

### ***Hechos\_pedido***

**ID\_hecho** INT PRIMARY KEY,  
**ID\_cliente** INT NOT NULL COMMENT 'ID del cliente asociado al pedido'

**ID\_producto** INT NOT NULL COMMENT 'ID del producto asociado al pedido'

**ID\_pago** INT NOT NULL COMMENT 'ID del pago asociado al pedido'

**ID\_tiempo** DATE NOT NULL COMMENT 'Fecha del pedido'

**cantidad** INT COMMENT 'Cantidad del producto en el pedido'

**precio\_unidad** DECIMAL(15,2) COMMENT 'Precio unitario del producto '

**total** DECIMAL(15,2) COMMENT 'Total del pedido'

### *Restricciones de integridad referencial*

FOREIGN KEY (ID\_cliente) REFERENCES dimension\_cliente (ID\_cliente),

FOREIGN KEY (ID\_producto) REFERENCES dimension\_producto (ID\_producto),

FOREIGN KEY (ID\_pago) REFERENCES dimension\_pago (ID\_pago),

FOREIGN KEY (ID\_tiempo) REFERENCES dimension\_tiempo (fecha\_pedido)

**FOREIGN KEY (ID\_cliente) REFERENCES dimension\_cliente (ID\_cliente):** Esta línea establece una relación de clave externa entre la columna ID\_cliente en la tabla hechos\_pedido y la columna ID\_cliente en la tabla dimension\_cliente. Esto garantiza que cada valor en ID\_cliente de hechos\_pedido debe tener una correspondencia en la tabla dimension\_cliente, lo que mantiene la integridad referencial entre ambas tablas.

**FOREIGN KEY (ID\_producto) REFERENCES dimension\_producto (ID\_producto):** Similar al anterior, esta línea establece una relación de clave externa entre la columna ID\_producto en la tabla hechos\_pedido y la columna ID\_producto en la tabla dimension\_producto. Esto asegura que cada valor en ID\_producto de hechos\_pedido esté asociado con un ID de producto válido en la tabla dimension\_producto.

**FOREIGN KEY (ID\_pago) REFERENCES dimension\_pago (ID\_pago):** Esta línea establece una relación de clave externa entre la columna ID\_pago en la tabla hechos\_pedido y la columna ID\_pago en la tabla dimension\_pago. De esta forma, cada ID de pago en hechos\_pedido debe tener una correspondencia en la tabla dimension\_pago.

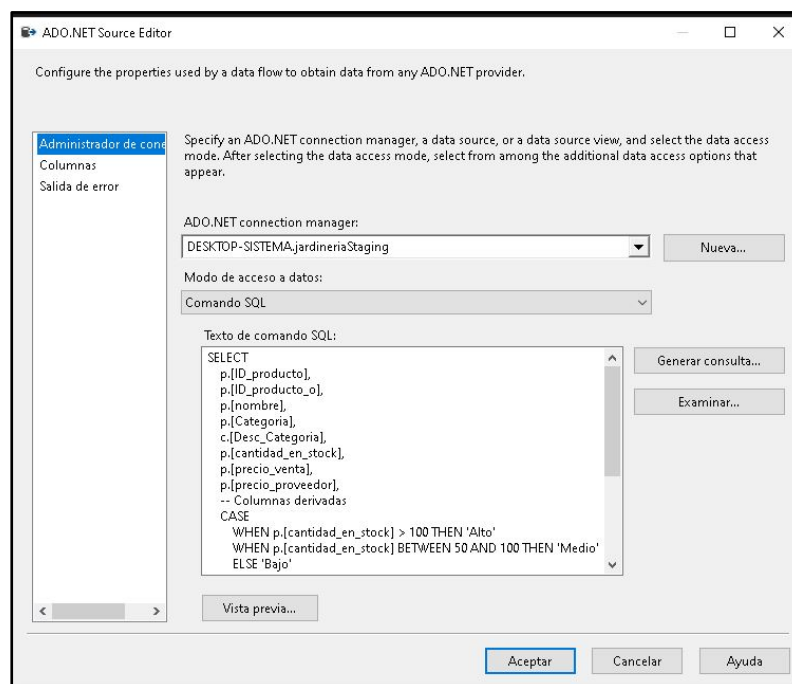
**FOREIGN KEY (ID\_tiempo) REFERENCES dimension\_tiempo (fecha\_pedido):** Aquí se establece una relación de clave externa entre la columna ID\_tiempo en hechos\_pedido y la columna fecha\_pedido en dimension\_tiempo. Esto asegura que cada fecha de pedido en hechos\_pedido esté asociada con una fecha válida en la tabla dimension\_tiempo.

## Proceso de transformación de datos y carga

### Origen de Datos (Staging)

#### 1. Configurar Origen de Datos (Staging):

- En SQL Server Management Studio (SSMS), selecciona el origen de datos donde se encuentran los datos iniciales.



- Configura una conexión de origen que apunte a la tabla "staging".
- Consultas Utilizadas:  
[https://drive.google.com/file/d/1rk\\_WDMxhgbhy1hZLr\\_NNAyyRSF\\_-fx-W/view?usp=drive\\_link](https://drive.google.com/file/d/1rk_WDMxhgbhy1hZLr_NNAyyRSF_-fx-W/view?usp=drive_link)
- Se crearon dos transformaciones cliente y Tiempo por medio de una derived column

## Tiempo

Derived Column Name	Derived Column	Expression
cliente_pais	<add as new column>	nombre_cliente + " (" + pais + ")"
tipo-localidad	<add as new column>	pais == "Spain" ? "Local" : "International"
tel_organizado	<add as new column>	REPLACE(REPLACE(REPLACE(telefono, " ", "" ), "-", "" ), "", "...")

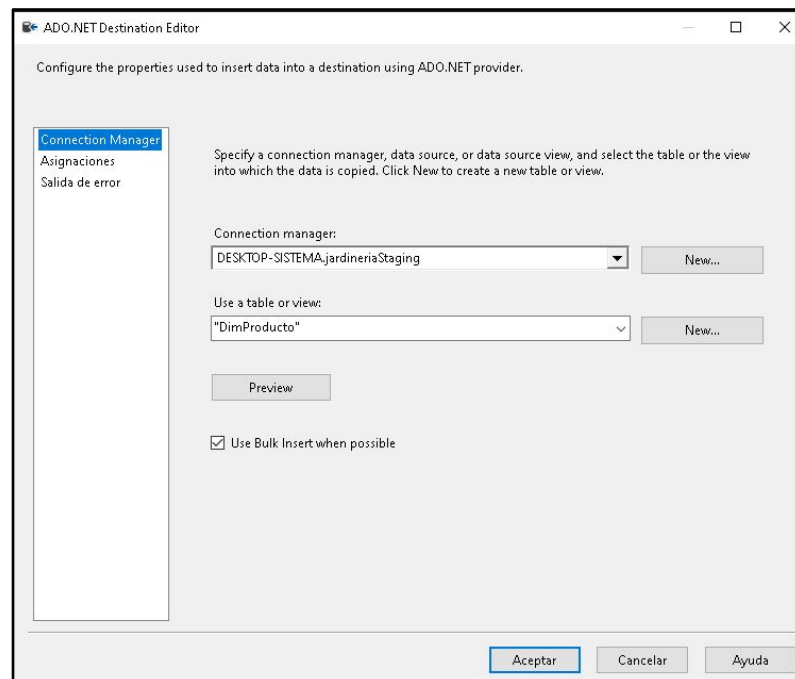
## Cliente

Año	<add as new column>	YEAR(fecha)
Mes	<add as new column>	MONTH(fecha)
Día	<add as new column>	DAY(fecha)
DíaSemanaNum	<add as new column>	DATEPART("dw", fecha)
Trimestre	<add as new column>	DATEPART("quarter", fecha)
Díaaño	<add as new column>	DATEPART("dy", fecha)

## Destino Dim Tabla

### 2. Configurar Destino de Datos (Dim):

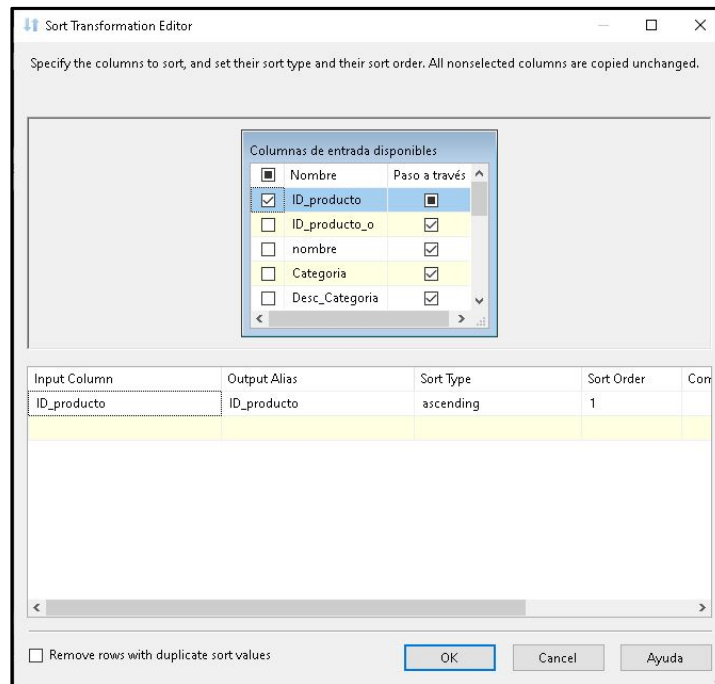
- Configura un destino de datos que apunte a la tabla de destino "dim", que es donde se cargarán los datos transformados.



## Sort 1 (Ordenar Datos del Origen)

### 3. Agregar Componente de Sort (Ordenar Datos del Origen):

- Añade un componente de "Sort" después del origen de datos. Este componente se utiliza para ordenar los datos según los criterios necesarios para las operaciones posteriores.

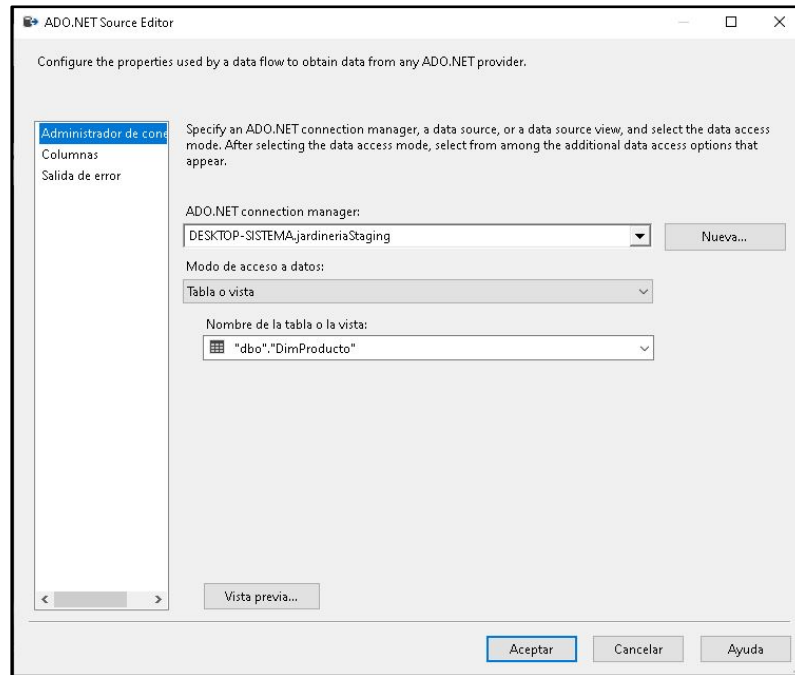


## Origen de Datos (DestinoDim)

### 4. Configurar Otro Origen de Datos (DestinoDim):

- Configura otro origen de datos que apunte a la tabla "DestinoDim". Este origen se utilizará para comparar las inserciones con los datos que ya existen en el destino.

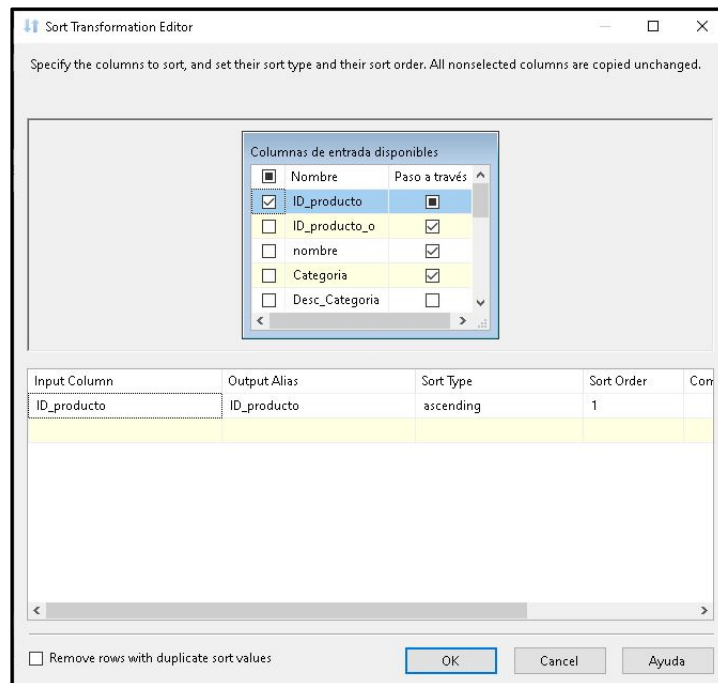




## Sort 2 (Ordenar Datos del Destino)

### 5. Agregar Segundo Componente de Sort (Ordenar Datos del Destino):

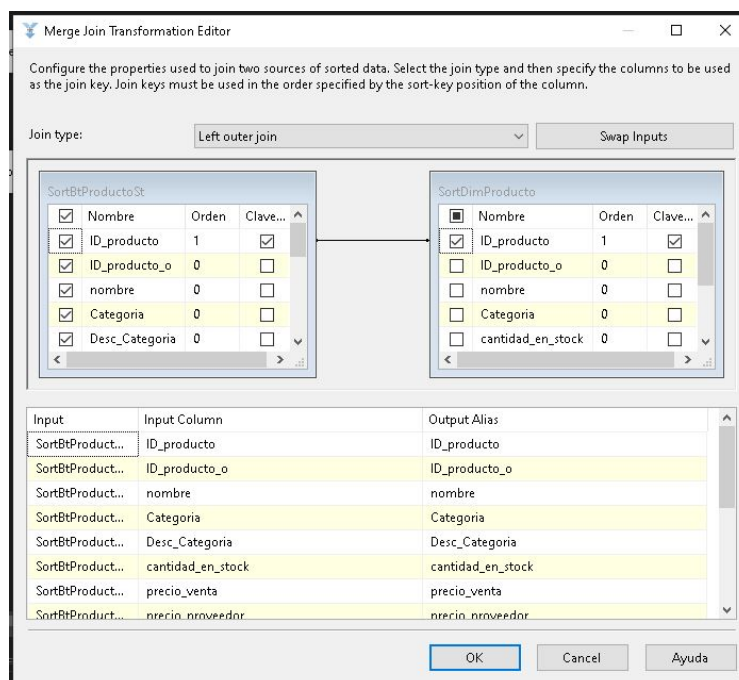
- Añade un segundo componente de "Sort" después del origen de datos "DestinoDim". Este componente ordenará los datos del destino para que coincidan con los datos del origen.



## Merge Join (Combinar Datos)

### 6. Utilizar Componente de Merge Join:

- Utiliza un componente de "Merge Join" para combinar los datos ordenados del origen y del destino.

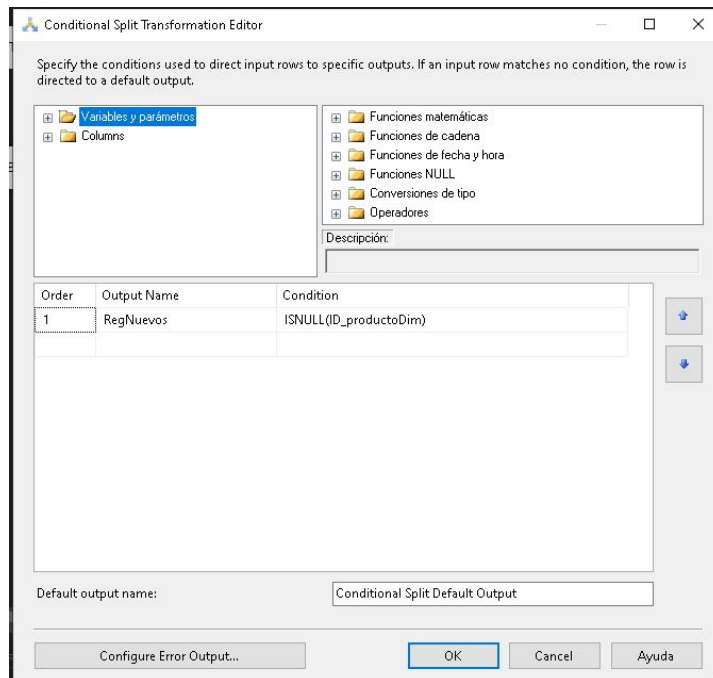


- El "Merge Join" se basa en una clave común, como una columna de identificación única, para combinar correctamente los datos.

## Conditional Split (Dividir Datos)

### 7. Agregar Componente de Conditional Split:

- Añade un componente de "Conditional Split" después del "Merge Join".

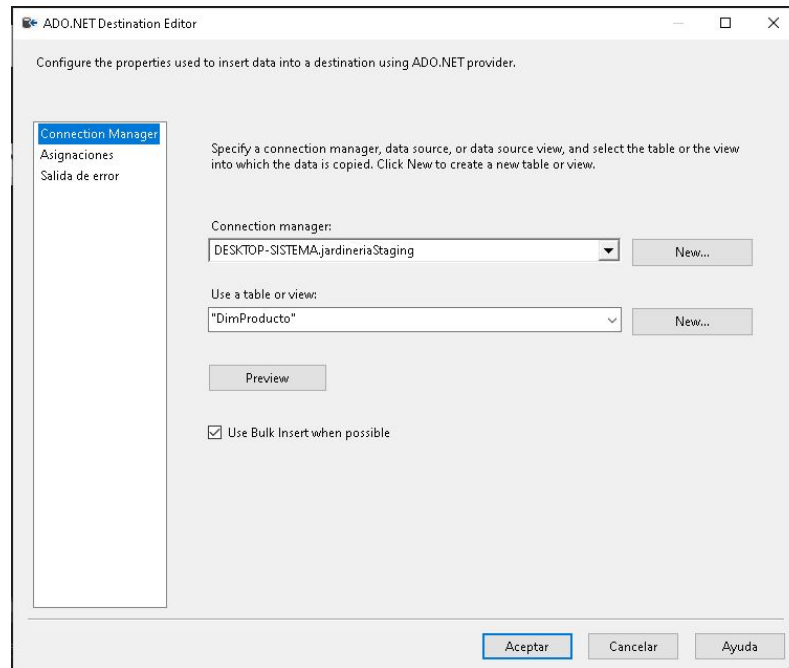


- Este componente se usa para dividir los datos según condiciones específicas. Por ejemplo, puede identificar nuevas inserciones, actualizaciones o registros existentes.

## Destino Dim Tabla (Resultados del Merge Join)

### 8. Conectar al Destino de Datos (DestinoDim):

- Conecta la salida del "Conditional Split" al destino de datos "DestinoDim".



- Este paso asegura que los resultados del proceso de combinación y división se carguen correctamente en la tabla de destino.

## **Configuración del Proceso de Carga Final en el Data Mart**

El proceso para la carga final en el datamart definitivo sigue siendo el mismo, con la diferencia de que el destino ahora es una base de datos nueva para el datamart final, mientras que las dimensiones creadas durante la transformación ahora son el origen.

### **Origen de Datos (Dimensiones Transformadas)**

#### **1. Configurar Origen de Datos (Dimensiones Transformadas):**

- Configura una conexión de origen que apunte a las tablas de dimensiones transformadas en tu base de datos intermedia o de staging. Estas tablas contienen los datos ya transformados y listos para ser cargados en el Data Mart final.

### **Destino de Datos (Data Mart Final)**

#### **2. Configurar Destino de Datos (Data Mart Final):**

- Configura una conexión de destino que apunte a la nueva base de datos del Data Mart final. Aquí es donde deseas cargar los datos completamente transformados.

### **Ordenar Datos del Origen**

#### **3. Agregar Componente de Sort (Ordenar Datos del Origen):**

- Añade un componente de "Sort" después del origen de datos para ordenar los datos según los criterios necesarios para las operaciones posteriores.

### **Origen de Datos (Datos Actuales del Data Mart)**

#### **4. Configurar Origen de Datos (Datos Actuales del Data Mart):**

- Configura otro origen de datos que apunte a las tablas del Data Mart final. Este origen se utilizará para comparar las inserciones y actualizaciones con los datos que ya existen en el Data Mart.

### **Ordenar Datos del Destino**

#### **5. Agregar Segundo Componente de Sort (Ordenar Datos del Destino):**

- Añade un segundo componente de "Sort" después del origen de datos del Data Mart. Este componente ordenará los datos del destino para que coincidan con los datos del origen.

#### **Combinar Datos (Merge Join)**

#### **6. Utilizar Componente de Merge Join:**

- Utiliza un componente de "Merge Join" para combinar los datos ordenados del origen y del destino.
- El "Merge Join" se basa en una clave común, como una columna de identificación única, para combinar correctamente los datos.

#### **Dividir Datos (Conditional Split)**

#### **7. Agregar Componente de Conditional Split:**

- Añade un componente de "Conditional Split" después del "Merge Join".
- Este componente se usa para dividir los datos según condiciones específicas, como identificar nuevas inserciones, actualizaciones o registros existentes.

#### **Cargar Datos en el Data Mart Final**

#### **8. Conectar al Destino de Datos (Data Mart Final):**

- Conecta la salida del "Conditional Split" al destino de datos en la nueva base de datos del Data Mart final.
- Asegura que los resultados del proceso de combinación y división se carguen correctamente en las tablas del Data Mart.

## Conclusiones

La implementación del modelo estrella en la base de datos de la empresa de jardinería, ha demostrado ser altamente efectiva para mejorar la eficiencia en el análisis de datos; debido a que su estructura multidimensional facilita la realización de consultas complejas y el análisis detallado de información, lo que permite a la empresa obtener insights valiosos sobre sus operaciones y clientes.

Gracias al modelo estrella, se ha observado una mejora significativa en la capacidad de la empresa para tomar decisiones estratégicas basadas en datos precisos y actualizados, gracias a la integración de dimensiones como tiempo, cliente, producto y pago se ha proporcionado una visión holística de las transacciones comerciales, lo que ha permitido identificar tendencias, patrones y oportunidades de mejora.

El diseño del modelo estrella, ha contribuido a una mayor agilidad en la generación de informes y análisis ad hoc; además, la estructura optimizada de la base de datos ha reducido los tiempos de respuesta en la obtención de información, lo que ha mejorado la capacidad de la empresa para responder rápidamente a las necesidades del negocio y a los cambios en el mercado.

La implementación de procesos de transformación y carga de datos (ETL) en un modelo estrella ha demostrado ser una estrategia eficaz para optimizar la gestión de datos en el sector de la jardinería, estos procesos aseguran la precisión y consistencia de los datos, facilitando un acceso y análisis más eficientes. Como resultado, se mejora la capacidad de realizar análisis profundos y detallados, lo que proporciona insights valiosos y permite tomar decisiones estratégicas más informadas, este enfoque no solo optimiza la operación diaria,

sino que también impulsa la capacidad de respuesta y adaptación ante las necesidades del mercado, demostrando el potencial significativo del modelo estrella en la gestión empresarial.



## Referencias

- García, J. (2013). *Diseño data warehouse: hechos y dimensiones; modelo estrella vs copo de nieve*. Obtenido de mundodb: <https://mundodb.es/disenio-data-warehouse-hechos-y-dimensiones-modelo-estrella-vs-copo-de-nieve>
- Rodriguez Afaro, F. (2019). *El modelo de estrella. El pilar fundamental del Business Intelligence*. Obtenido de datamanagement: <https://datamanagement.es/2019/06/27/business-intelligence-modelo-estrella/>
- Sánchez, J., & Canelo , C. (2017). *Modelo de data warehouse con aplicacion de inteligencia de negocios para las pymes*. Obtenido de Universidad Nacional Jorge Basadre Grohmann: <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwi-zryUsuaFAxWNSDABHShGDvU4RhAWegQIDxAB&url=https%3A%2F%2Frevistas.unjbg.edu.pe%2Findex.php%2Fcyd%2Farticle%2Fdownload%2F737%2F749%2F1384&usg=AOvVaw0On9M4FBKuUhy1Avd1W>
- Zerpa , H., García , R., & Izquierdo , H. (2020). *Datamart basado en el modelo estrella para la implementación de indicadores clave de desempeño como salida del big data*. Obtenido de Universidad Nacional Experimental Politécnica: [https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjt9KHtsOaFAxUGSDABHd-sBtAQFnoECDcQAQ&url=https%3A%2F%2Ftunexpo.autanabooks.com%2Findex.php%2Ftun%2Farticle%2Fdownload%2F342%2F618%2F&usg=AOvVaw0xm28ke\\_Kas10lQrvxcXl](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjt9KHtsOaFAxUGSDABHd-sBtAQFnoECDcQAQ&url=https%3A%2F%2Ftunexpo.autanabooks.com%2Findex.php%2Ftun%2Farticle%2Fdownload%2F342%2F618%2F&usg=AOvVaw0xm28ke_Kas10lQrvxcXl)