

**Materia: Electiva II Bases de datos II**

**Docente: Victor Hugo Mercado**

## **Módulo 3**

**Proceso de transformación de  
datos y carga en el data mart final**



**Grupo 5:**

- **Carlos Andrés Bustamante Montoya**
- **Santiago Osorio Henao**

**Especialización en Analítica y Big Data**

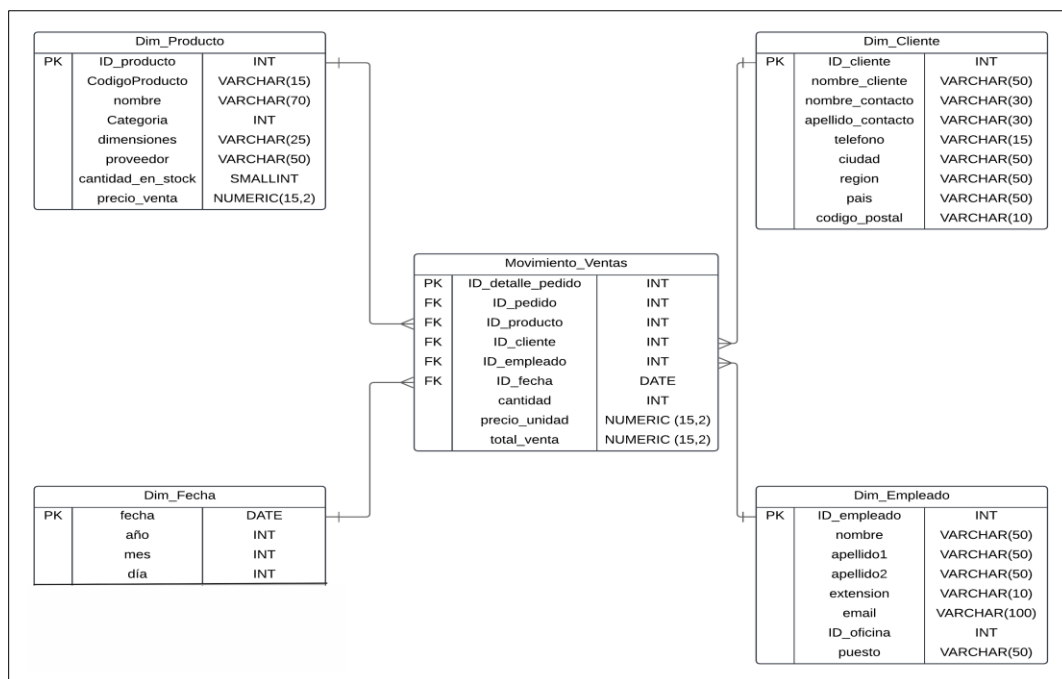
Medellín, 09 de Diciembre de 2024

## Transformación y carga de datos de una base de datos de origen al data mart final para el Análisis de Ventas en la Empresa “Jardinería”

### Descripción de modelo de estrella propuesto

El modelo estrella para "Jardinería" está centrado en la tabla de hechos llamada **“Movimiento\_Ventas”**, que representa cada transacción de venta de productos. Esta tabla se relaciona con dimensiones que describen los detalles contextuales de cada venta, como el producto, el cliente, el empleado que gestionó la venta y la fecha en que se realizó la transacción. Cada dimensión almacena detalles específicos que permiten analizar las ventas desde diferentes perspectivas.

### Diseño del modelo estrella



### Lista de dimensiones propuestas

#### 1. Dimensión de Producto (dim\_producto)

**Descripción:** Proporciona detalles de cada producto vendido.

**Columnas y tipos de datos:**

- **ID\_producto (INT)** - Llave primaria.
- **CodigoProducto (VARCHAR(15))** - Código único del producto.
- **nombre (VARCHAR(70))** - Nombre descriptivo del producto.
- **Categoria (INT)** - Identificador de la categoría del producto.
- **dimensiones (VARCHAR(25))** - Dimensiones del producto.
- **proveedor (VARCHAR(50))** - Nombre del proveedor del producto.

- **cantidad\_en\_stock (SMALLINT)** - Cantidad disponible en inventario.
- **precio\_venta (NUMERIC(15,2))** - Precio de venta del producto.

## 2. Dimensión de Cliente (dim\_cliente)

**Descripción:** Contiene información sobre los clientes que realizan las compras.

**Columnas y tipos de datos:**

- **ID\_cliente (INT)** - Llave primaria.
- **nombre\_cliente (VARCHAR(50))** - Nombre de la empresa o cliente.
- **nombre\_contacto (VARCHAR(30))** - Nombre del contacto principal.
- **apellido\_contacto (VARCHAR(30))** - Apellido del contacto.
- **telefono (VARCHAR(15))** - Teléfono de contacto.
- **ciudad (VARCHAR(50))** - Ciudad de residencia.
- **region (VARCHAR(50))** - Región de residencia.
- **pais (VARCHAR(50))** - País de residencia.
- **codigo\_postal (VARCHAR(10))** - Código postal.

## 3. Dimensión de Empleado (dim\_empleado)

**Descripción:** Registra información sobre los empleados que gestionan las ventas.

**Columnas y tipos de datos:**

- **ID\_empleado (INT)** - Llave primaria.
- **nombre (VARCHAR(50))** - Nombre del empleado.
- **apellido1 (VARCHAR(50))** - Primer apellido.
- **apellido2 (VARCHAR(50))** - Segundo apellido.
- **extension (VARCHAR(10))** - Extensión telefónica.
- **email (VARCHAR(100))** - Correo electrónico.
- **ID\_oficina (INT)** - Identificador de la oficina.
- **puesto (VARCHAR(50))** - Puesto del empleado.

## 4. Dimensión de Fecha (dim\_fecha)

**Descripción:** Permite análisis mediante el desglose de la fecha de venta.

**Columnas y tipos de datos:**

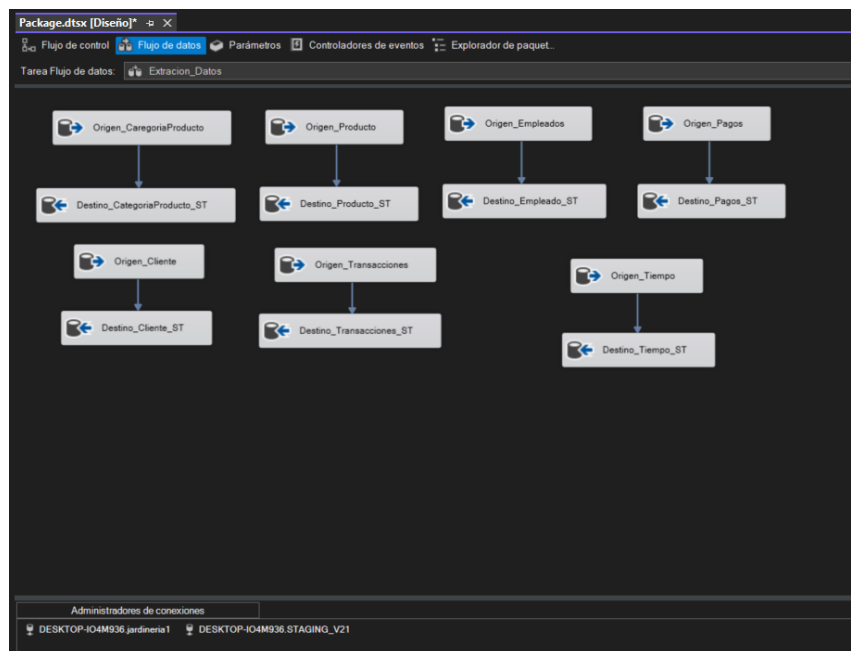
- **fecha (DATE)** - Llave primaria.
- **año (INT)** - Año de la fecha.
- **mes (INT)** - Mes de la fecha.
- **día (INT)** - Día de la fecha.

## Tabla Hechos

La tabla de hechos “**Movimiento\_Ventas**” almacena las transacciones de venta

- **ID\_detalle\_pedido (INT (PK))**- Identificador único de la transacción de venta.
- **ID\_pedido (INT)**- Relación con la tabla de pedidos.
- **ID\_producto (INT)**- Identificador del producto vendido.
- **ID\_cliente (INT)**- Identificador del cliente.
- **ID\_empleado (INT)**- Identificador del empleado de ventas.
- **ID\_fecha (DATE)**- Fecha de la venta.
- **Cantidad (INT)**- Cantidad de producto vendida.
- **precio\_unidad (NUMERIC(15,2))**- Precio de venta unitario.
- **total\_linea (NUMERIC(15,2))**- Total de la línea (cantidad \* precio\_unidad).

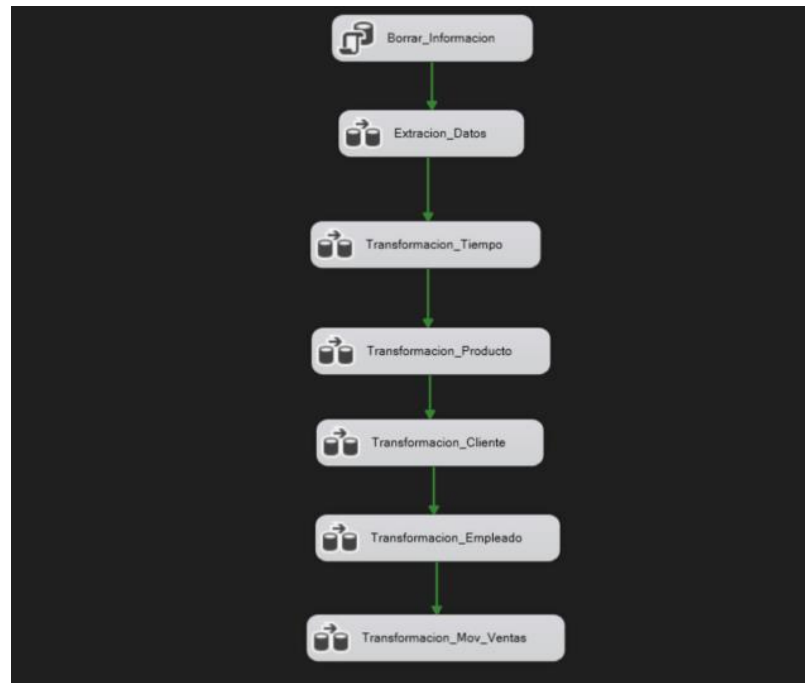
En las anteriores actividades se había hecho la creación de las bases de datos de origen y Staging, además de realizar el proceso de extracción.



Continuaremos entonces con el siguiente paso que es la transformación y validación de los datos:

### 1-Pasos de Transformación:

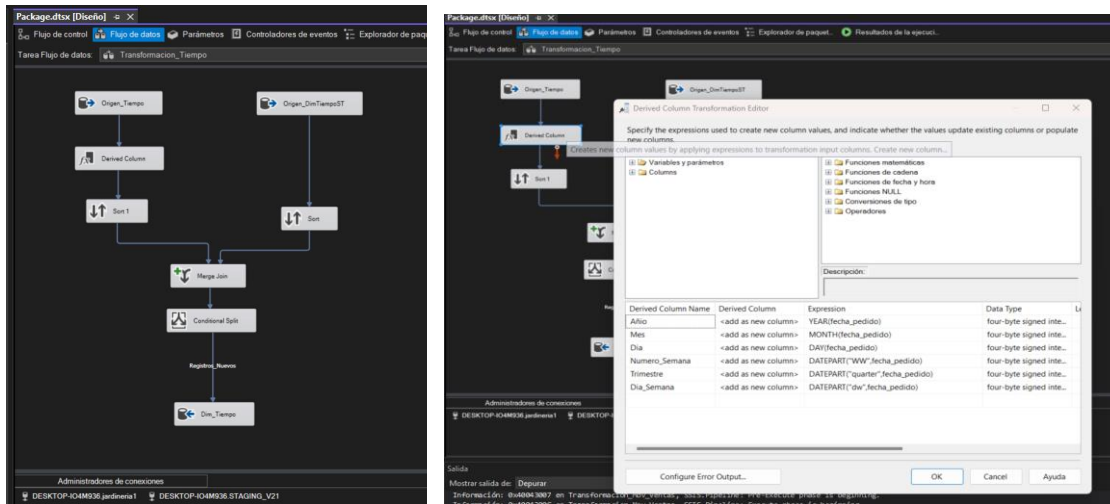
Se realiza la creación de “**Data Flow Task**” por cada uno de las dimensiones para realizar las transformaciones de la siguiente forma:



En cada uno de las transformaciones se debe tener en cuenta un origen, el cual será la tabla que se creó previamente en la evidencia anterior cuando se hizo la extracción y un destino que será el espacio de la nueva dimensión ya con la limpieza y transformación de datos:

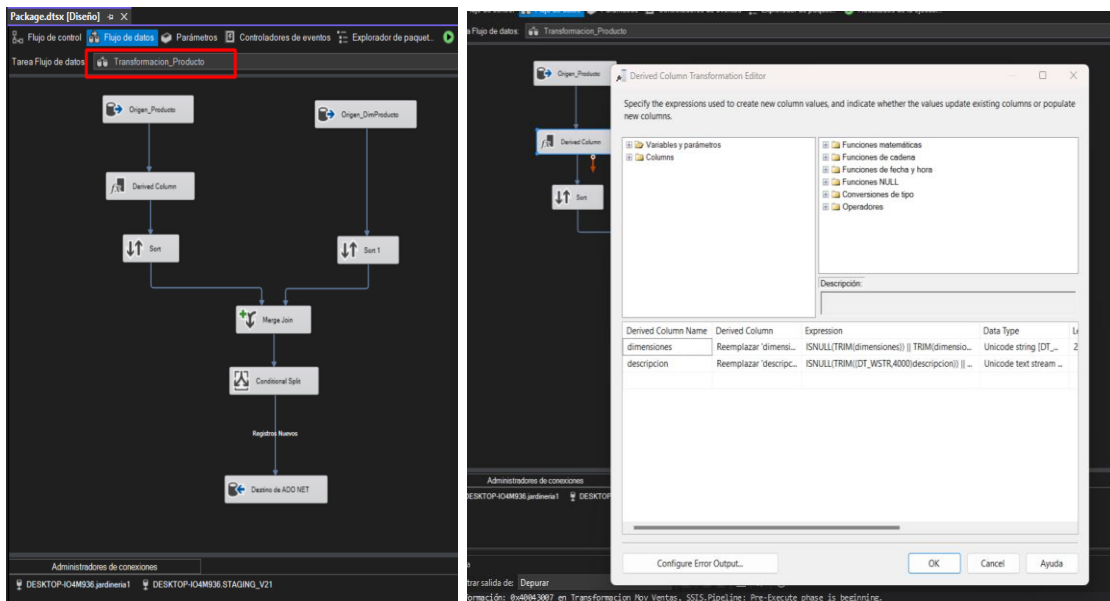
## 2-Creacion y Transformación de Dimensiones

### -Dimensión Tiempo:

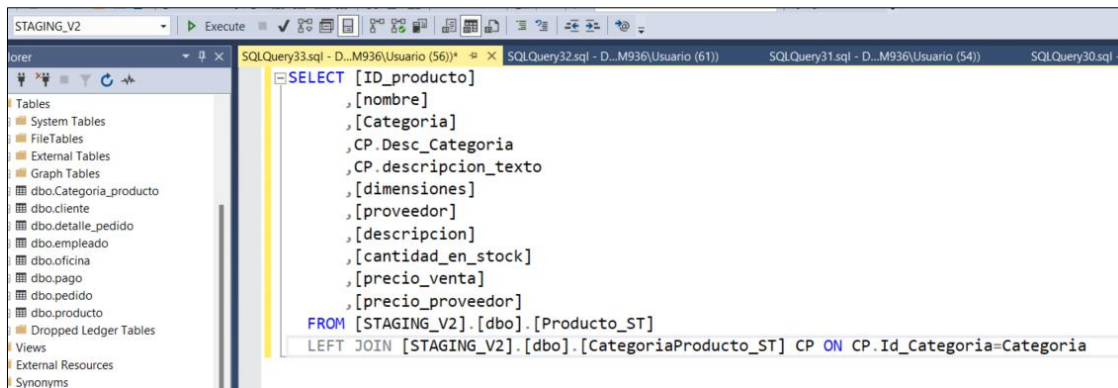


Dentro de las transformaciones se realizó la creación de varias columnas las cuales con ayuda de las funciones de fecha y hora pudimos crear, el año, mes, día, numero\_semana, trimestre, Dia\_semana de acuerdo al campo “fecha\_pedido”, además de hacer una ordenación de columnas y agregar la función **condicional Split** que nos ayudará a evitar registros duplicados basándose en los resultados de la comparación en el componente “**Merge Join**” durante la carga de información.

### -Transformación Producto:

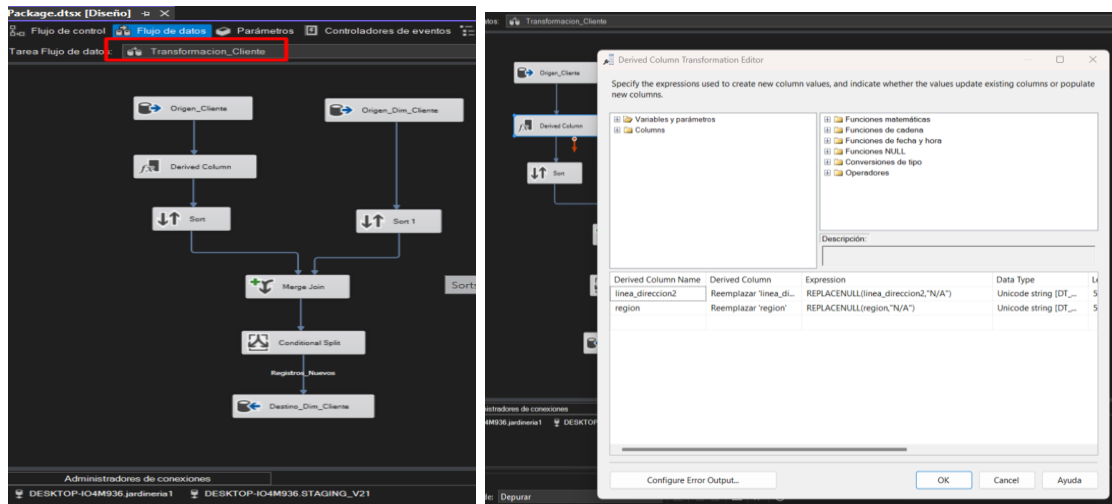


Para el origen de la dimensión se creo desde SQL ya que necesitábamos información de otra tabla para complementar y que sea más estándar la información:



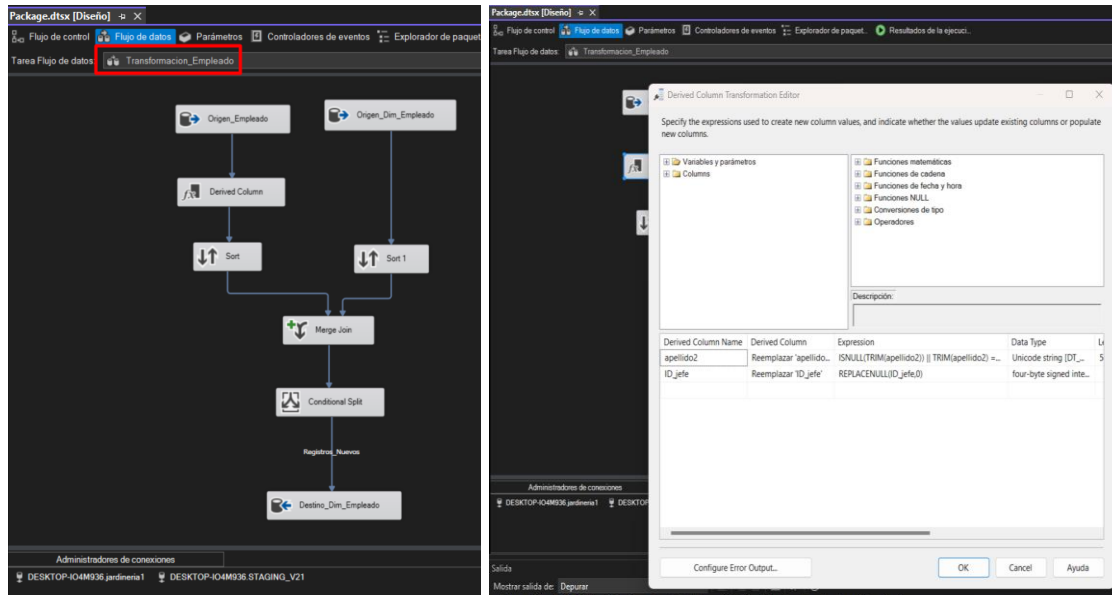
Dentro de este paso se realiza la transformación de las columnas dimensiones y descripción ya que estas tenían registros en NULL y se reemplazan por “N/A”, al igual que el anterior paso se utiliza el componente “Merge Join” y “Condicional Split” para evitar la carga de duplicados

### -Transformación Cliente:



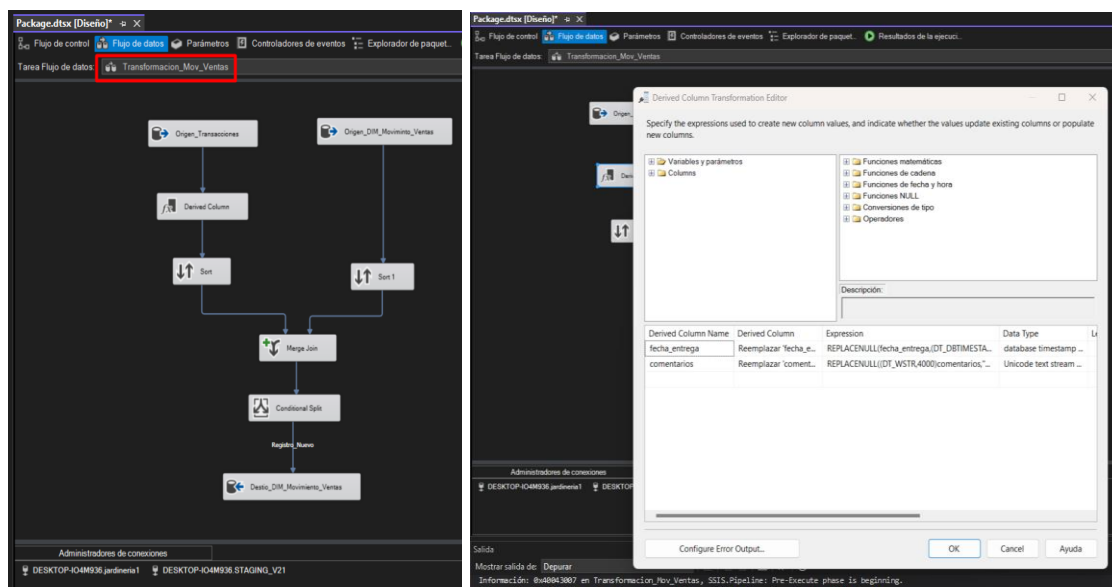
Se realiza transformaciones a las columnas “línea\_direccion2” y “región” ya que en algunos registros contienen NULL y son sustituidos por “N/A”, además se realiza una validación para no cargar datos duplicados.

## -Transformación Empleado:



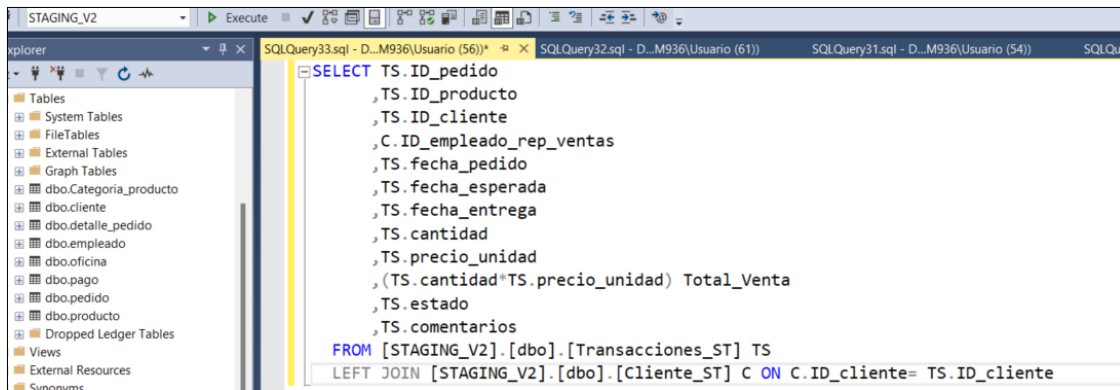
Dentro del campo “apellido1” contada con registros en blanco sin ninguna información, todos estos fueron reemplazados por “N/A”, además en ID\_jefe existía un registro con NULL, este se reemplaza por el valor cero, y se agrega los componentes para evitar la carga de registros duplicados.

## -Transformación Movimiento\_Ventas:



Para el origen de esta dimensión se utilizaron campos de otras tablas para, esta estructura se creó en SQL de esta forma:





Con esto tendríamos una dimensión con mayor información y más completa, se realizó varias transformaciones una en el campo “fecha\_entrega” ya que tenía registros con NULL y el campo “comentarios” de igual forma, y su validación con ayuda de componentes para evitar cargar datos duplicados.

### 3-Resultado:

Como resultado de todo el tratamiento de datos de datos nos queda la siguientes tablas o dimensiones en nuestra base de datos STAGING, sin ningún registro vacío, null o dato extraño:

