

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

León Felipe Sánchez Palacio

Duvan Quintero Guerra

Institución Universitaria Digital de Antioquia

Bases de Datos II (PREICA2402B010070)

Victor Hugo Mercado

06 de octubre de 2024

Introducción

El modelado tipo estrella es una estructura que nos va a permitir optimizar el desempeño al momento de hacer consultas y al mismo tiempo poder leer estas. Este nos va a permitir ejecutar y analizar de una manera más eficaz, por eso en este trabajo buscaremos mostrar el cómo podemos pasar de un modelo entidad relación a un modelo tipo estrella y los beneficios que este nos da al momento de hacer el análisis a los diferentes datos que estemos manejando.

Planteamiento del problema

Se requiere construir un modelo estrella el cual tenga tabla desechos y tablas de dimensiones que permitan analizar de una forma eficiente las ventas y las transacciones que realiza una empresa que se dedica a la venta de artículos de jardinería.

Análisis del problema

Se tiene la necesidad de construir un modelo estrella para la empresa que se dedica a la venta de artículos de jardinería. El objetivo que se tiene es diseñar una estructura de base de datos que nos permita el análisis de datos como saber cual es el producto más vendido, la categoría que cuenta con más productos y cual es el año que tiene más ventas.

El componente clave en este caso es lograr que la relación entre las diferentes dimensiones, con la tabla de hechos se den de una forma que se logre optimizar las diferentes consultas y así lograr un análisis más a fondo, que nos brinden una mejor toma de decisiones para la empresa.

Propuesta de la solución

Se diseñará un modelo estrella que permitirá analizar de manera eficiente las ventas y transacciones de la empresa. Este modelo incluirá una tabla de hechos de pedidos, que integrará las dimensiones propuestas: clientes, empleados, productos, oficinas y tiempo. De esta manera, se facilitará una evaluación detallada de los distintos datos a estudiar.

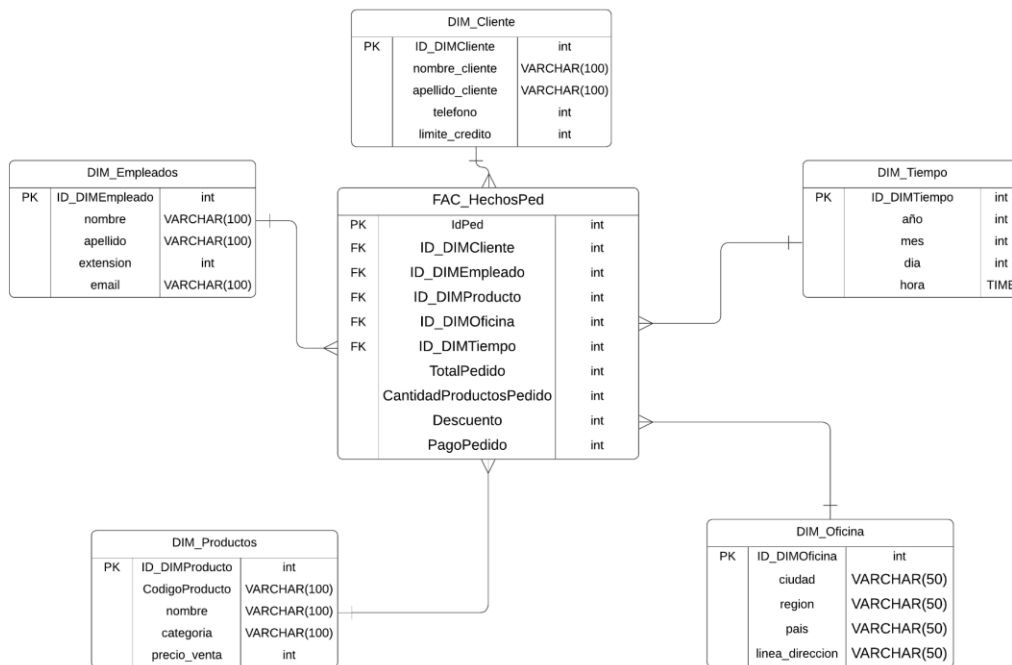


Tabla de hechos FAC_HechosPed

ID_DIMCliente

ID_DIMEmpleado

ID_DIMProducto

ID_DIMOficina

ID_DIMTiempo

TotalPedido

CantidadProductosPedido

Descuento

PagoPedido

DIM_Cliente

ID_DIMCliente INTEGER identity(1,1) NOT NULL **PRIMARY KEY**,

nombre_cliente VARCHAR(30) DEFAULT NULL,

apellido_cliente VARCHAR(30) DEFAULT NULL,

telefono VARCHAR(15) NOT NULL,

limite_credito NUMERIC(15,2) DEFAULT NULL

DIM_Empleado

ID_DIMEmpleado INT identity(1,1) NOT NULL **PRIMARY KEY**,

nombre VARCHAR(50) NOT NULL,

apellido VARCHAR(50) NOT NULL,

extension VARCHAR(10) NOT NULL,

email VARCHAR(100) NOT NULL

DIM_Producto

ID_DIMProducto INTEGER identity(1,1) NOT NULL **PRIMARY KEY**,

CodigoProducto VARCHAR(15) NOT NULL,
nombre VARCHAR(70) NOT NULL,
Categoria int NOT NULL,
precio_venta NUMERIC(15,2) NOT NULL

DIM_Tiempo

ID_DIMTiempo

Año

Mes

Dia

Hora

DIM_Oficina

ID_DIMOficina

ciudad VARCHAR(50) NOT NULL,

region VARCHAR(50) DEFAULT NULL,

pais VARCHAR(50) DEFAULT NULL,

linea_direccion VARCHAR(50) NOT NULL

Descripción del análisis realizado

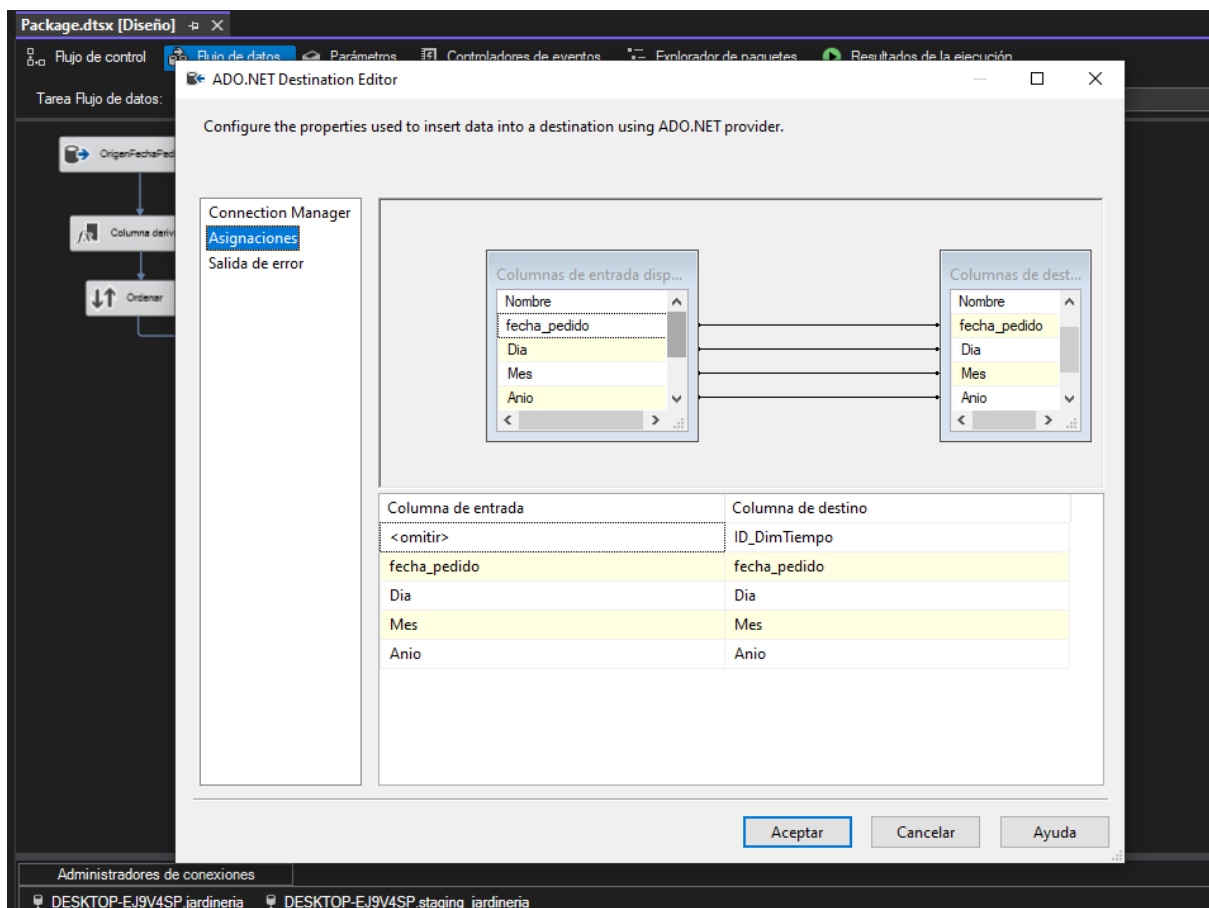
Para hacer el respectivo traslado de los datos los cuales se usaran en la base de datos Stating, se uso el modelo estrella realizado en la primer entrega, esto ayudándonos a hacer realizar el traslado de una manera optima y de una manera eficaz, gracias a esto de la tabla detalle pedido, hemos traído el id, y la cantidad de productos, de la tabla pedido, hemos traído el id del pedido y el id cliente, los cuales nos ayudaran para consultar datos en nuestra FAC de hechos, también de la tabla pedido, hemos traigo las fechas de pedido y fecha de entrega. De la tabla producto hemos añadido el id del producto, el código del producto, nombre, categoría y precio de venta, estos datos nos ayudaran para hacer consultas de acuerdo con el producto o su categoría. Siguiendo con la tabla de pago, hemos traído los datos, de id pago, id cliente, fecha pago y el total, así con estos datos podremos consultar cual cliente realizo el pago, también en que fecha se hizo el pago y el total pagado de alguno de los pedidos. De la tabla cliente hemos trasladado las variables de id cliente, teléfono, nombre, dirección, ciudad, país y limite, con estos tendremos los datos básicos del cliente, pero a su vez tendremos el lugar geográfico de cada uno, permitiéndonos realizar consultas como de qué país o ciudad se realiza una mayor cantidad de pedidos entre otros. Tenemos la tabla empleada de está hemos trasladado los datos básicos como lo son nombre y apellido, pero también hemos traído datos como id oficina y id jefe, lo cual nos permitirá saber el empleado que hizo x pedido a que oficina pertenece, adicional trasladamos datos de contacto como email y extensión, el cual hace referencia al número de teléfono. Y por último de la tabla oficina hemos trasladado el id oficina, país, ciudad, región y dirección, estos permitiéndonos hacer consultas de ubicación de las diferentes sedes de la jardinería.

proceso de transformación de datos

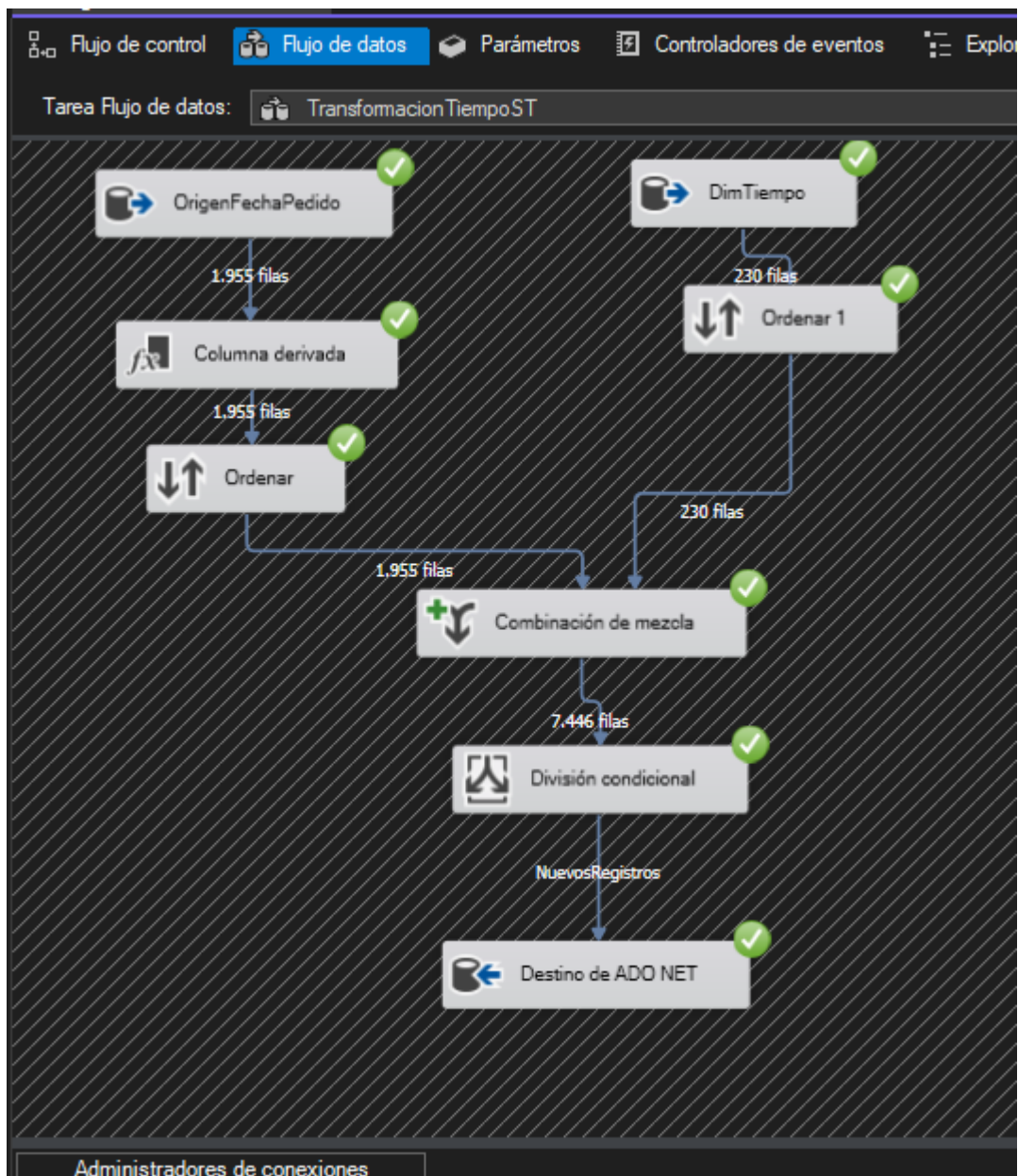
Para hacer nuestro proceso de transformación de datos, nos basamos en las tablas que se extrajeron en las actividades anteriores, y en nuestro modelo estrella realizado también anteriormente. Estos pasos han sido fundamentales, ya que nos permitirán realizar el proceso de transformación y posteriormente el de carga de una manera exitosa y optima.

A continuación, se mostrará el proceso realizado:

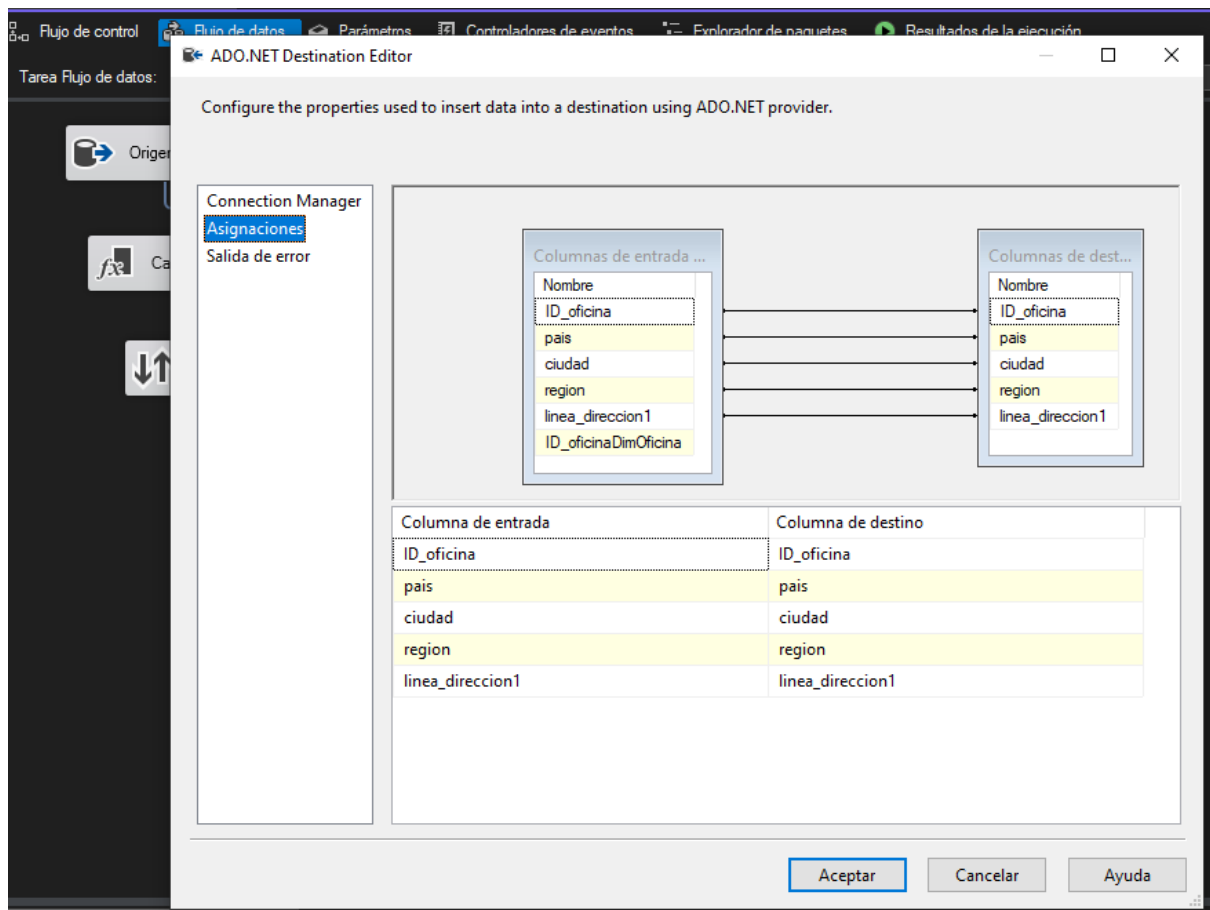
Iniciamos con nuestros datos para la dimensión de tiempo, el cual consta de la fecha de pedido, y esta se discrimina por día, mes y año:



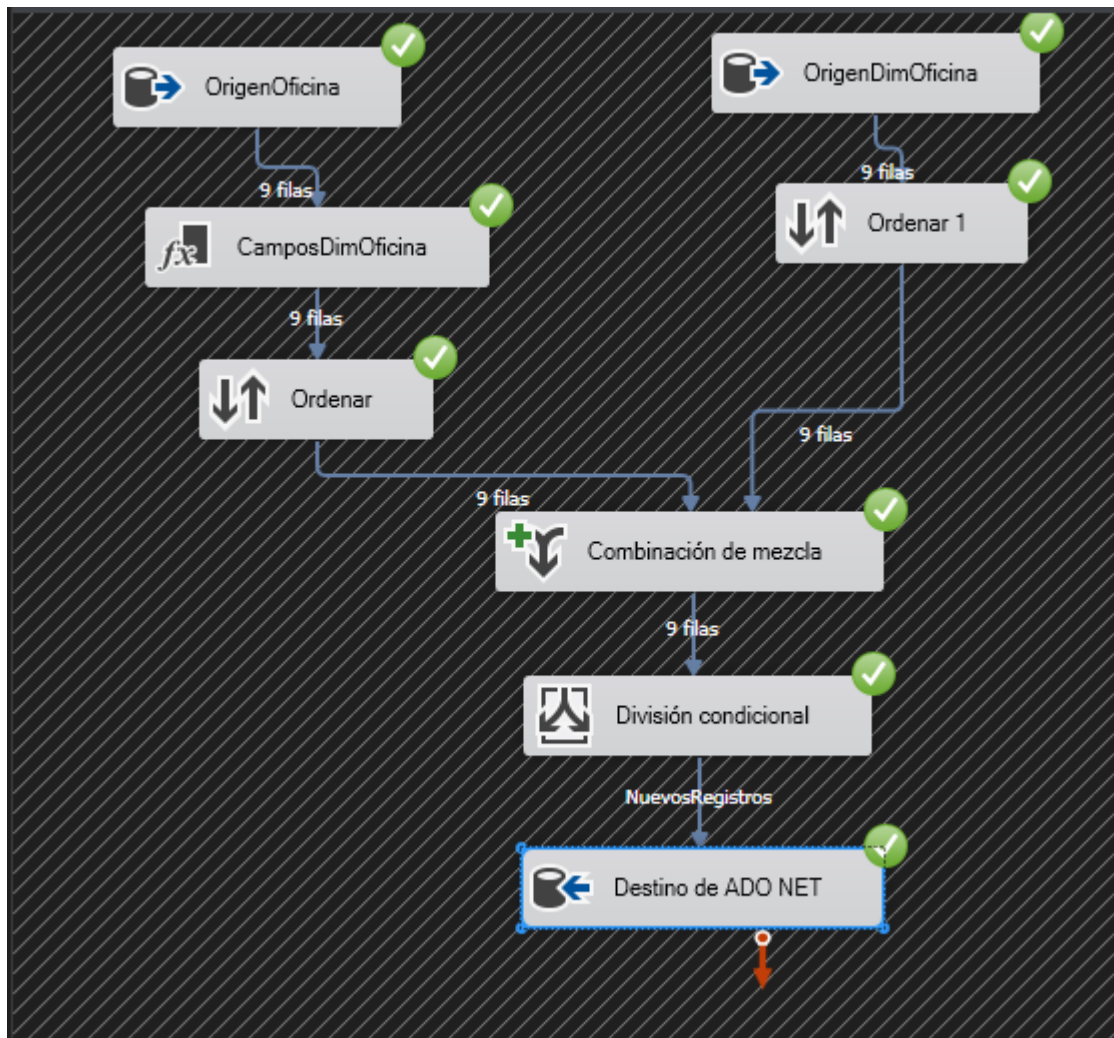
Y al ejecutar nuestra transformación, observamos que como anteriormente ya se habían cargado los datos, esta no cargara de nuevo datos repetidos:



Nuestra segunda transformación fue la de los datos de la dimensión oficina:



Esta dimensión es la que nos va a permitir traer los datos geográficos para nuestro data mart



La tercer transformación realizada es la de los datos para nuestra dimensión de productos, la cual nos va a permitir saber el detalle que tendrá cada pedido.

Configure the properties used to insert data into a destination using ADO.NET provider.

Connection Manager

Asignaciones

Salida de error

Columnas de entrada dispo...

Nombre
ID_producto
CodigoProducto
nombre
Categoria
precio_venta
ID_productoDimProducto

Columnas de destin...

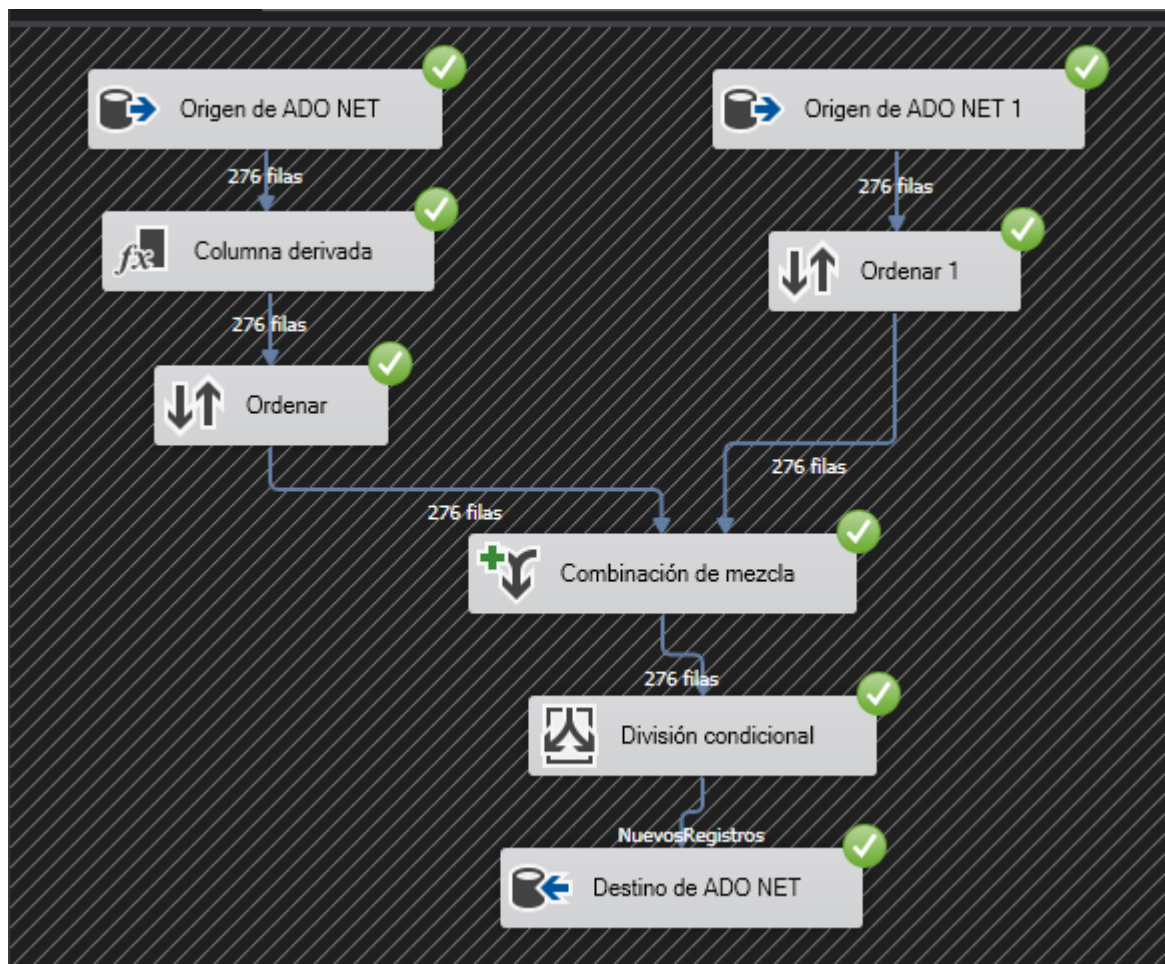
Nombre
ID_producto
CodigoProducto
nombre
Categoria
precio_venta

Columna de entrada	Columna de destino
ID_producto	ID_producto
CodigoProducto	CodigoProducto
nombre	nombre
Categoria	Categoria
precio_venta	precio_venta

Aceptar

Cancelar

Ayuda



La penúltima transformación de datos realizada fue la de la dimensión empleados, esto con el fin de poder conocer la persona que realizó las y también saber la oficina desde que se realizó alguna venta.

Configure the properties used to insert data into a destination using ADO.NET provider.

Connection Manager

Asignaciones

Salida de error

Columnas de entra...

Nombre

ID_empleado

nombre

apellido1

apellido2

extension

email

ID_oficina

ID_jefe

Columnas de dest...

Nombre

ID_empleado

nombre

apellido1

apellido2

extension

email

ID_oficina

ID_jefe

Columna de entrada

ID_empleado

nombre

apellido1

apellido2

extension

email

ID_oficina

ID_jefe

Columna de destino

ID_empleado

nombre

apellido1

apellido2

extension

email

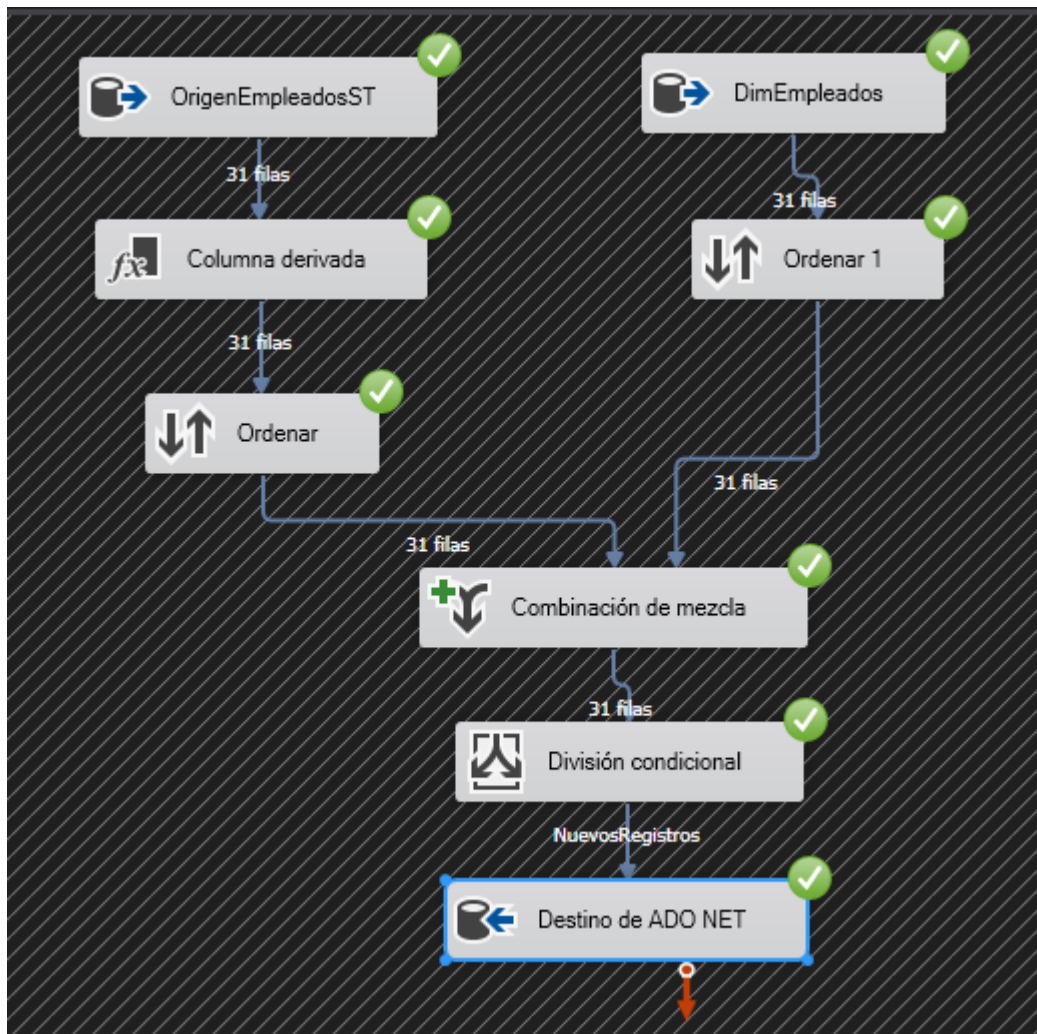
ID_oficina

ID_jefe

Aceptar

Cancelar

Ayuda



Y por ultimo tenemos los datos para nuestra dimensión de cliente, con esta podremos saber el nombre de la persona que hizo alguna compra, también el país, y la región desde donde se hizo, e igualmente el crédito que estas tienen.

Configure the properties used to insert data into a destination using ADO.NET provider.

Connection Manager
Asignaciones
Salida de error

Columnas de entra...

Nombre
ID_cliente
telefono
nombre_cliente
linea_direccion1
ciudad
region
pais

Columnas de desti...

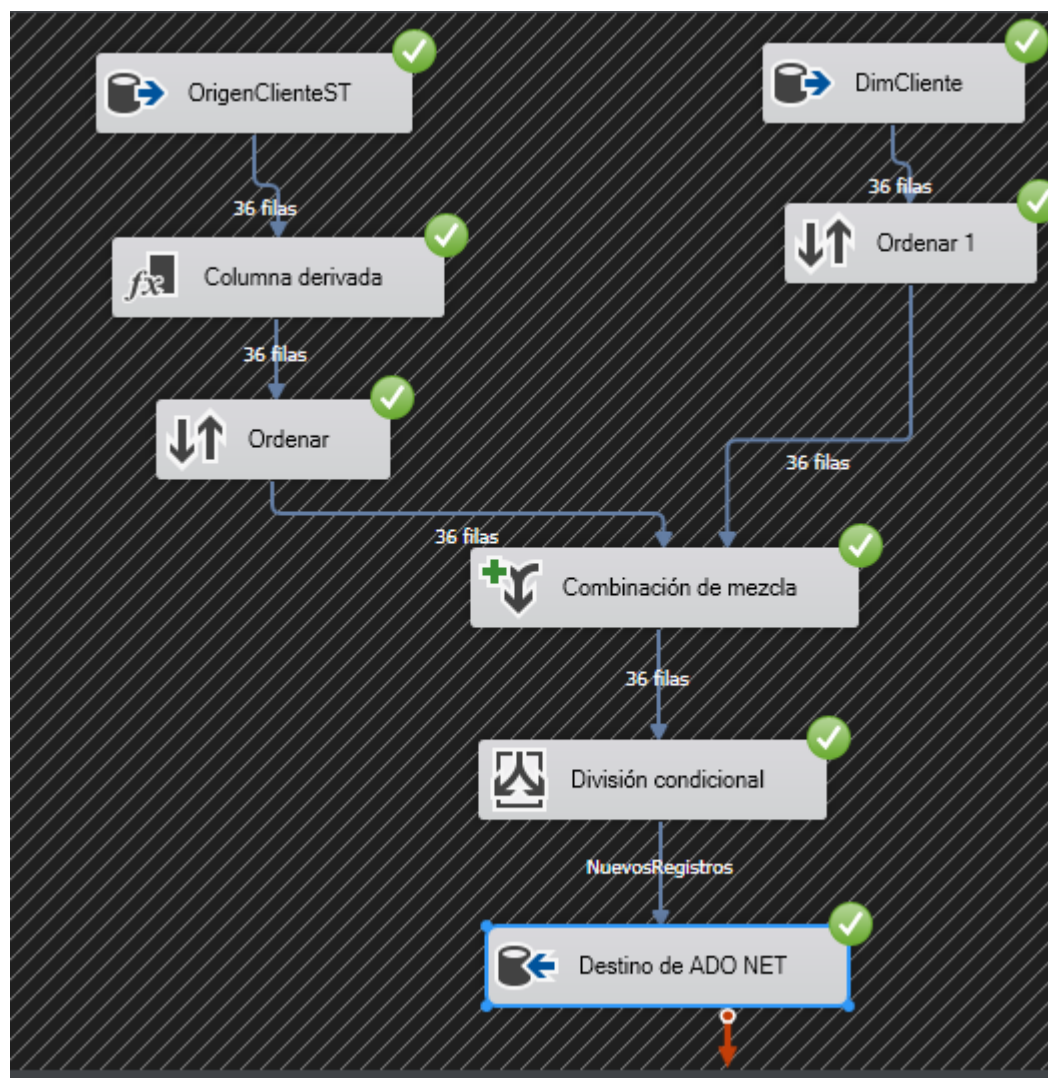
Nombre
ID_cliente
telefono
nombre_cliente
linea_direccion1
ciudad
region
pais

Columna de entrada	Columna de destino
ID_cliente	ID_cliente
telefono	telefono
nombre_cliente	nombre_cliente
linea_direccion1	linea_direccion1
ciudad	ciudad
region	region
pais	pais
limite_credito	limite_credito

Aceptar

Cancelar

Ayuda



Conclusiones

El modelo estrella es una herramienta que nos va a optimizar de una manera muy eficiente el análisis de datos, este va a mejorar el rendimiento de las consultas y va a facilitar la toma de decisiones. Este implementado de una manera adecuada va a mejorar de una manera significativa la capacidad que tenga cualquier tipo de empresa o entidad al momento de tomar decisiones que estén basadas en los diferentes tipos de datos que se generen dentro de ella.

Al momento de hacer el diseño de las estructuras de las tablas que irán en nuestra base de stating, podemos observar que hay muchas variables las cuales de una forma u otra son un poco innecesarias, el modelo estrella nos ha permitido optimizar la cantidad de tablas y variables que necesitamos de una base de datos para poder realizar las diferentes consultas de datos que se nos soliciten, esto sin duda alguna es una gran manera de agilizar y de hacer mas eficiente las diferentes tareas que se nos asignen, y de igual manera nos ayudara a tener ideas mas claras al momento de tomar ciertas decisiones.

La transformación de los datos es un paso fundamental en este proceso que se está llevando con la base de datos jardinería, pues es en este paso donde nos encargamos de preparar nuestros datos para cargarlos en nuestro data mart, y es acá donde debemos garantizar que los datos tengan una buena calidad y que la coherencia de estos esté de la mejor manera posible. Estos procesos nos han mostrado la importante de hacer una depuración de una manera efectiva y del como esta nos va a permitir optimizar nuestros procesos.