EVIDENCIA DE APRENDIZAJE 2. CREACIÓN DE UNA BASE DE DATOS DE STAGING

VICTOR HUGO MERCADO - PROFESOR |

OSCAR DAVID MADRIGAL FONNEGRA

BASES DE DATOS II -IUD

2024

**INTRODUCCIÓN**

La gestión eficiente de los datos es crucial para mejorar la toma de decisiones, optimizar los procesos comerciales y satisfacer las necesidades de los clientes. Para lograrlo, es fundamental contar con una infraestructura de bases de datos bien diseñada que permita almacenar, procesar y analizar la información de manera efectiva. En este proyecto, nos enfocamos en mejorar la gestión de datos de la bd Jardinería, mediante la construcción de una base de datos Staging que facilite el proceso de Extracción, Transformación y Carga (ETL) de los datos provenientes de su base de datos principal.

**OBJETIVOS**

**OBJETIVO GENERAL:**

* Diseñar e implementar una base de datos Staging que permita la extracción de datos desde la base de datos principal de Jardinería, con el fin de facilitar el análisis y la toma de decisiones.

**OBJETIVOS ESPECÍFICOS:**

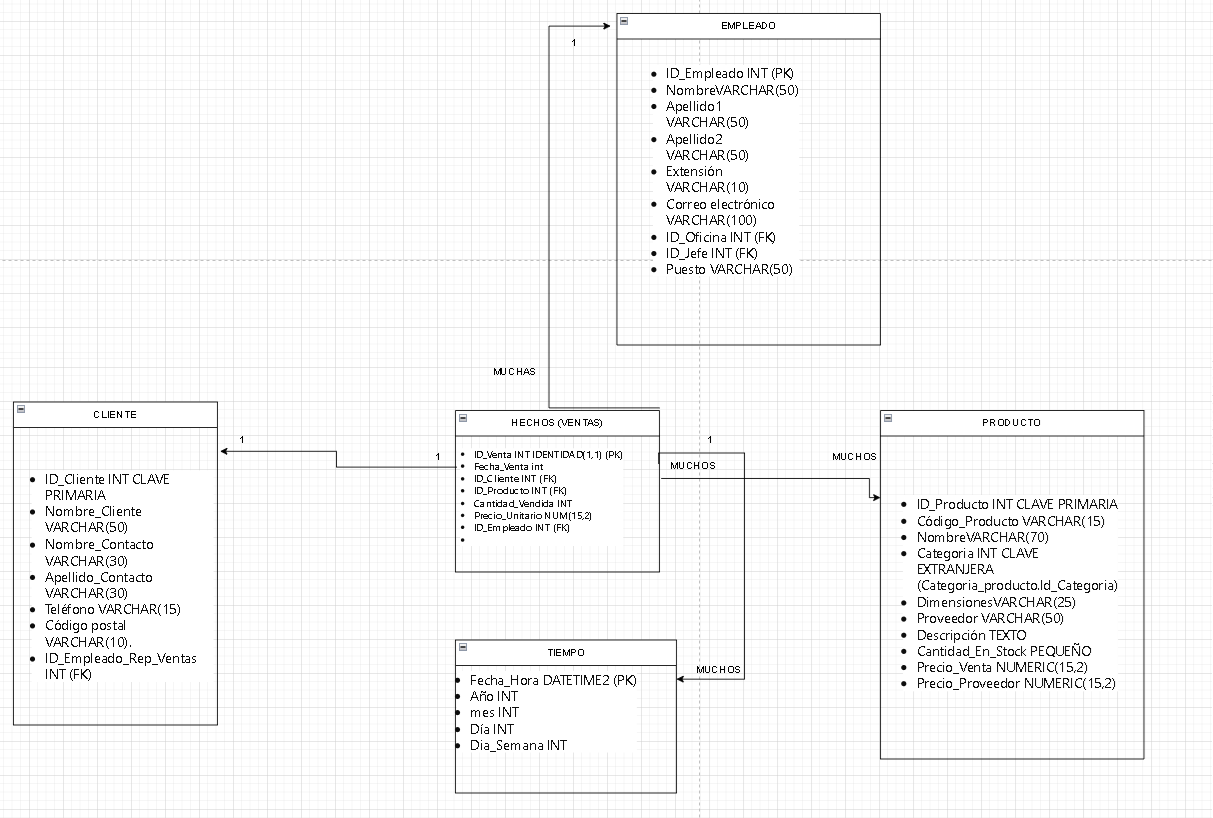
1. Realizar un análisis de los datos almacenados en la base de datos de Jardinería para identificar la información relevante y necesaria para el proceso de ETL.
2. Diseñar la estructura de las tablas que conformarán la base de datos Staging, teniendo en cuenta los requisitos de almacenamiento y las relaciones entre los datos.
3. Construir consultas SQL para extraer y cargar los datos desde la base de datos de Jardinería hacia la base de datos Staging.

**PLANTEAMIENTO DEL PROBLEMA**

La gestión de los datos en lo que respecta a la extracción, transformación y carga desde base de datos principal hacia entornos de análisis y reporting hace necesario crear una estructura adecuada para este proceso y así llevar acabo los análisis precisos y oportunos, lo que puede afectar la toma de decisiones y la eficiencia operativa de la empresa. Por lo tanto, es necesario diseñar e implementar una solución que permita optimizar el manejo de los datos y mejorar la capacidad de análisis de la empresa.

**ANÁLISIS DEL PROBLEMA**

El análisis de los datos almacenados en la base de datos de Jardinería revela la necesidad de una solución que facilite la extracción y transformación de estos datos para su posterior análisis. Se identifica la falta de una estructura centralizada y estandarizada para el almacenamiento de datos y se implementa en un formato de dimensiones, así como la ausencia de un proceso formalizado para la extracción de datos hacia entornos de análisis. Esto dificulta la integración de datos de diferentes fuentes, la realización de análisis históricos y la generación de informes precisos y oportunos. Además, se observa la presencia de datos redundantes y desactualizados, lo que afecta la integridad y la calidad de la información.

1. Análisis de los datos almacenados en la base de datos Jardinería con sus correcciones anteriores :
   * Revisar los datos almacenados en el Jardinería para identificar cuáles son relevantes y cuáles se deben trasladar a la base de datos *Staging.*
2. Construcción de la base de Datos *Staging:*
   * Diseñar la estructura tablas que estarán en la base de datos *Staging*.

Primero creamos la tabla de Hechos en en la base de datos orignen en sql server

CREATE TABLE Ventas (

ID\_Venta INT IDENTITY(1,1) PRIMARY KEY,

Fecha\_Venta DATE,

ID\_Cliente INT,

ID\_Producto INT,

Cantidad\_Vendida INT,

Precio\_Unitario NUMERIC(15, 2),

ID\_Empleado INT,

);

y creamos nuestra tabla tiempo

CREATE TABLE Tiempo (

Fecha DATE PRIMARY KEY,

Anio INT,

Mes INT,

Dia INT,

Dia\_Semana INT

);

Luego de esto vamos a crear nuestra stangin StaginJardineria, la cual vamos a utilizar para nuestra ETL en visual studio

Las consultas para dicho proceso son las siguientes donde realizaremos y trasladaremos los datos más relevantes

Tabla clientes:

SELECT id\_cliente,nombre\_cliente, nombre\_contacto, apellido\_contacto, telefono, linea\_direccion1 as Direccion, codigo\_postal, limite\_credito

FROM cliente;

Tabla producto:

SELECT ID\_producto, nombre, Categoria, dimensiones, proveedor, descripcion, cantidad\_en\_stock, precio\_venta, precio\_proveedor

FROM producto;

Tabla Ventas:

select ID\_venta,Fecha\_Venta,ID\_Cliente,ID\_Producto,Cantidad\_Vendida,Precio\_Unitario,ID\_Empleado

from Ventas

Tabla Tiempo:

select id\_tiempo, Anio,Mes,Dia,Dia\_Semana

from tiempo

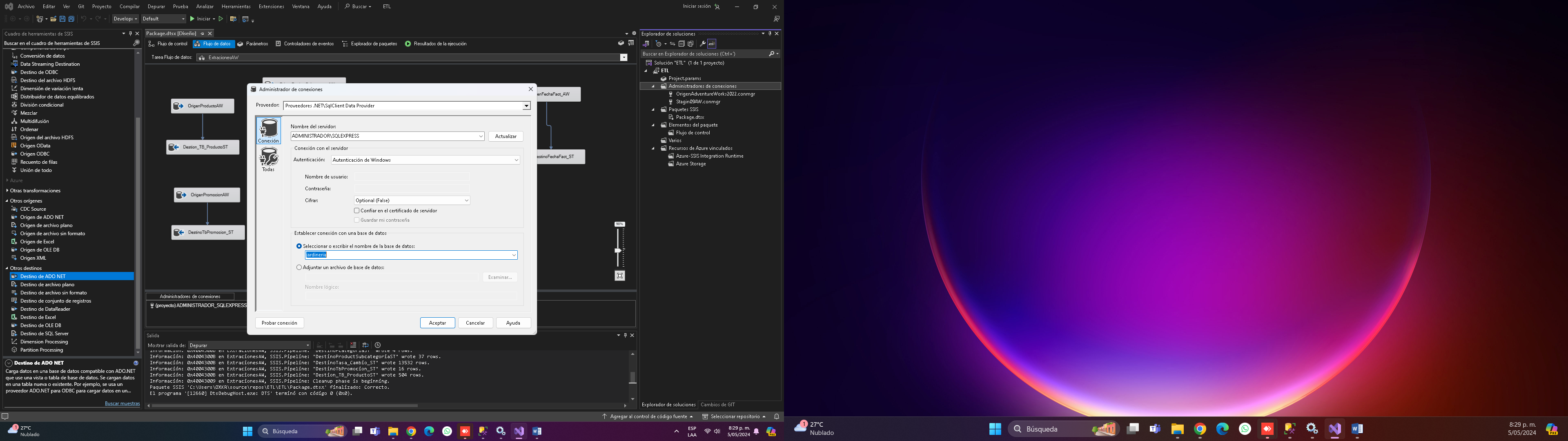
Tabla Empleado:

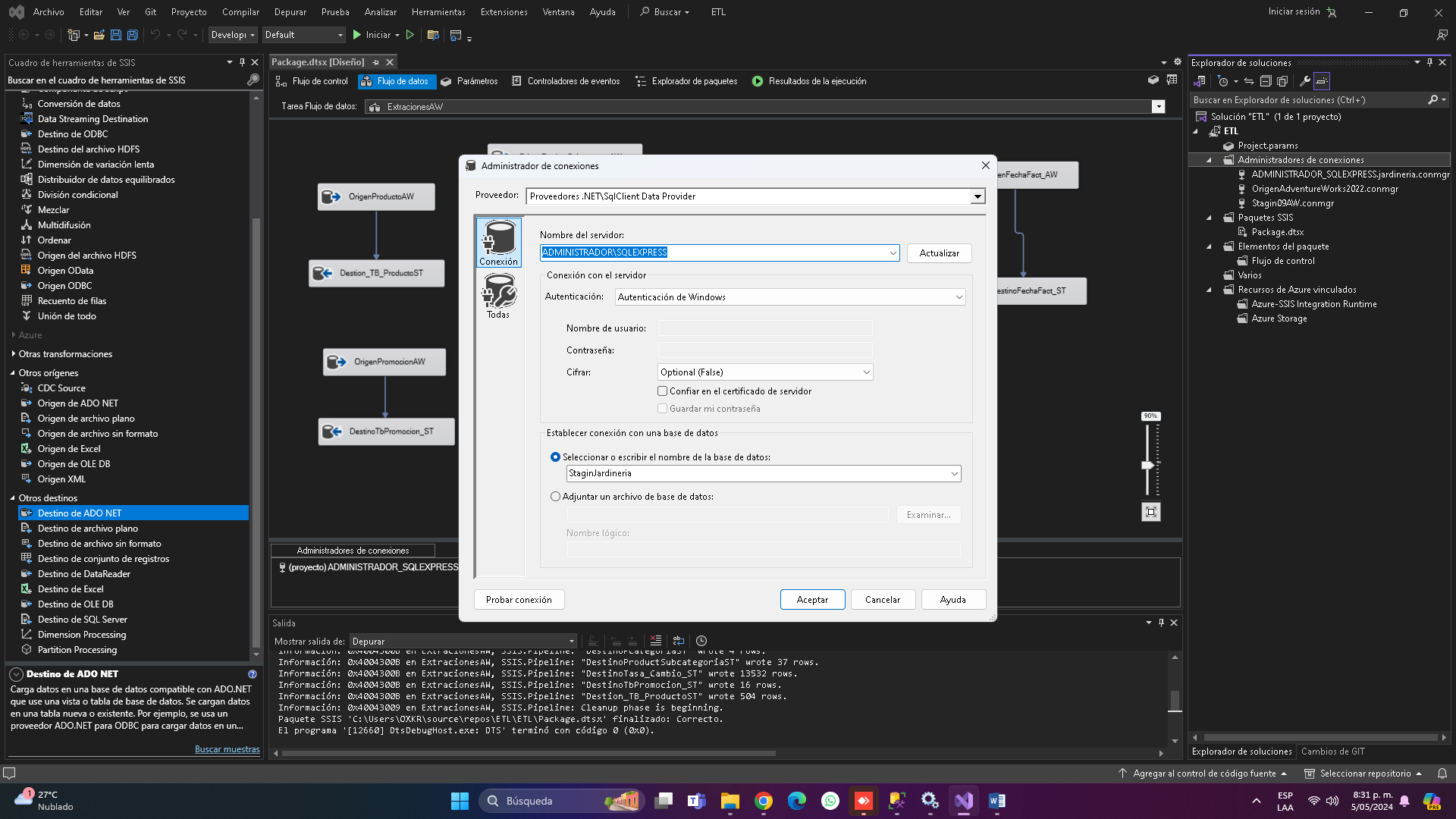
SELECT id\_empleado,nombre, apellido1, apellido2, ID\_oficina, ID\_jefe, puesto

FROM empleado;

3.Ahora creamos las conexiones en el Visual Studio para la base Jardinería –Origen

Y para el destino StaginJardineria





4. Ahora vamos a creamos el Flujo de Control y el Flujo de datos para cada una de las Tablas.

Vamos hacer un ejemplo solo por uno del origen al destino, y así se hace con cada una de las sentencias Sql de consultas de cada tabla, luego mostraremos cuando ejecutemos el Flujo completo en el Visual

CREATE TABLE "DestinoVentas\_ST" (

"ID\_venta" int identity(1,1),

"ID\_venta" int\_0,

"Fecha\_Venta" date,

"ID\_Cliente" int,

"ID\_Producto" int,

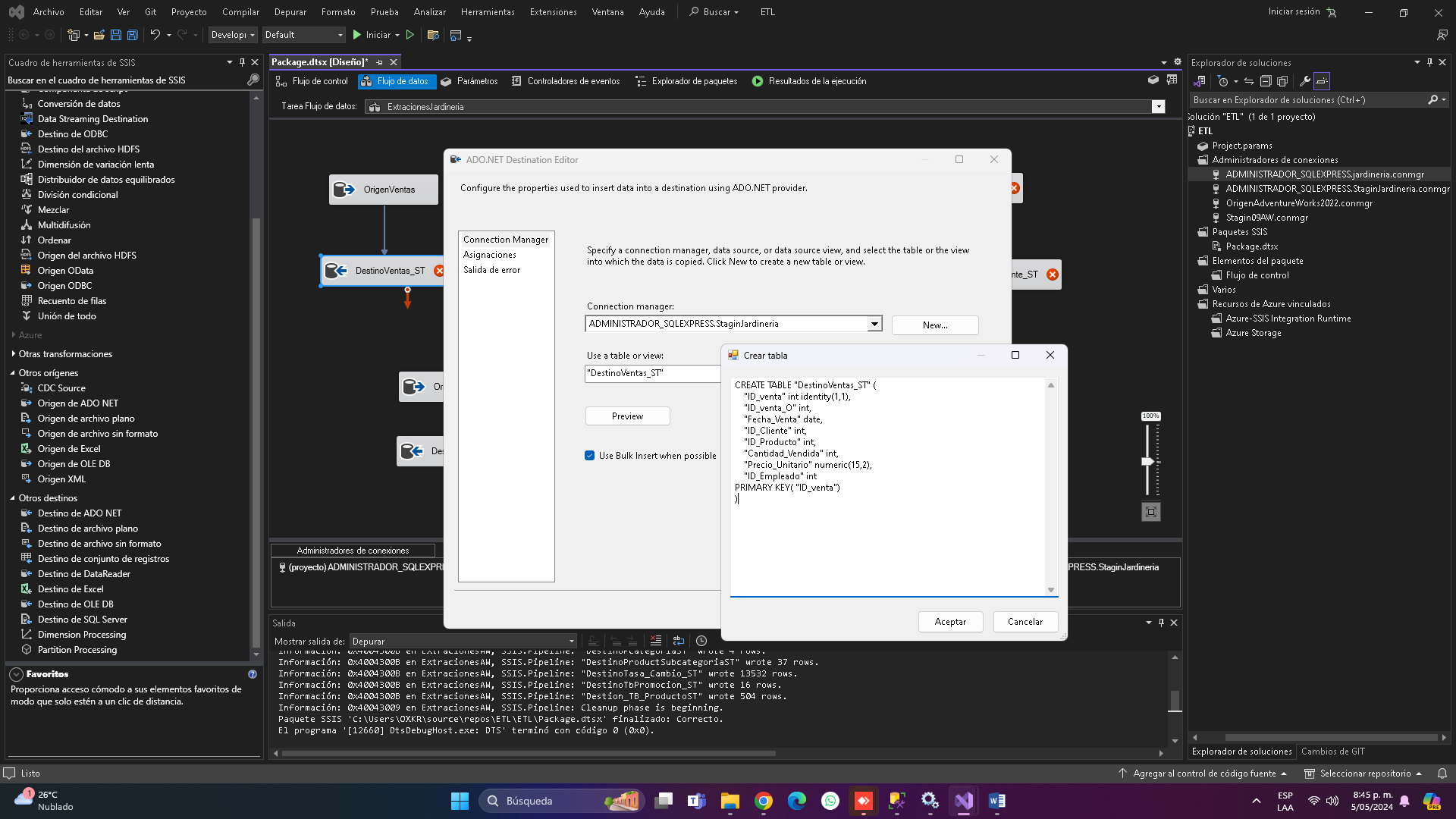
"Cantidad\_Vendida" int,

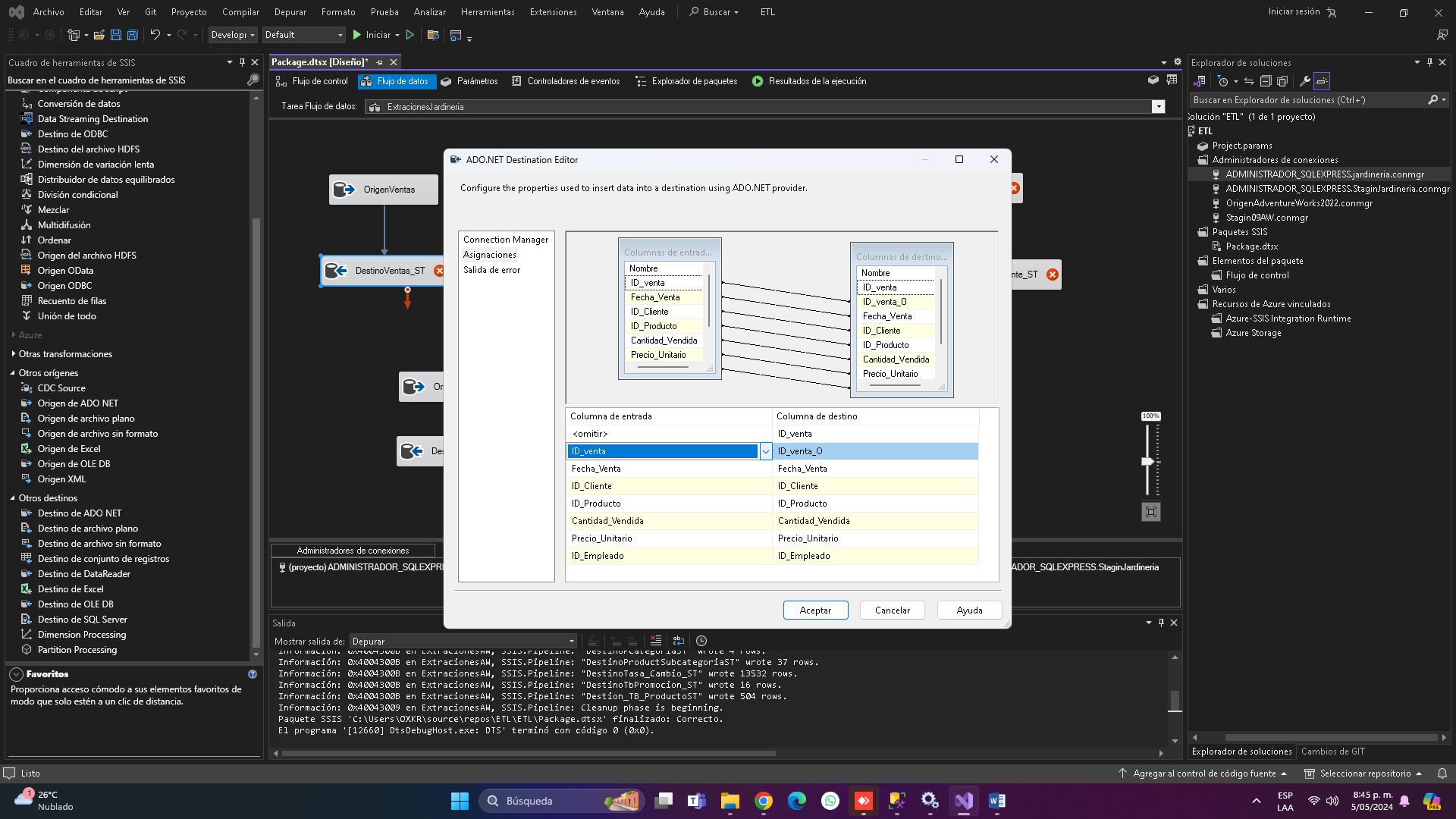
"Precio\_Unitario" numeric(15,2),

"ID\_Empleado" int,

Primary Key( "ID\_venta")

)





CREATE TABLE "DestinoEmpleado\_ST" (

"id\_empleado" int identity(1,1),

"id\_empleado\_0" int,

"nombre" nvarchar(50),

"apellido1" nvarchar(50),

"apellido2" nvarchar(50),

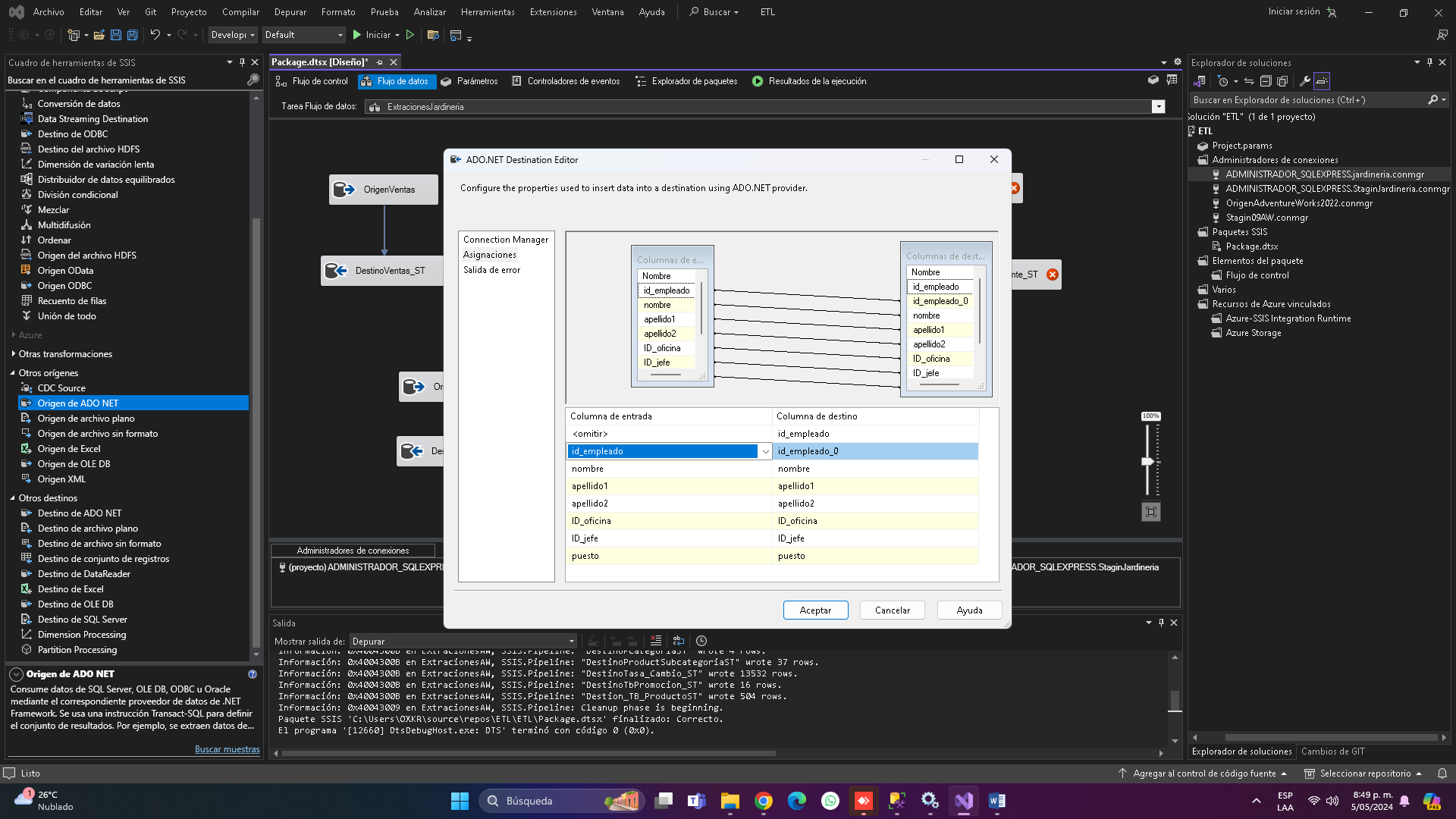
"ID\_oficina" int,

"ID\_jefe" int,

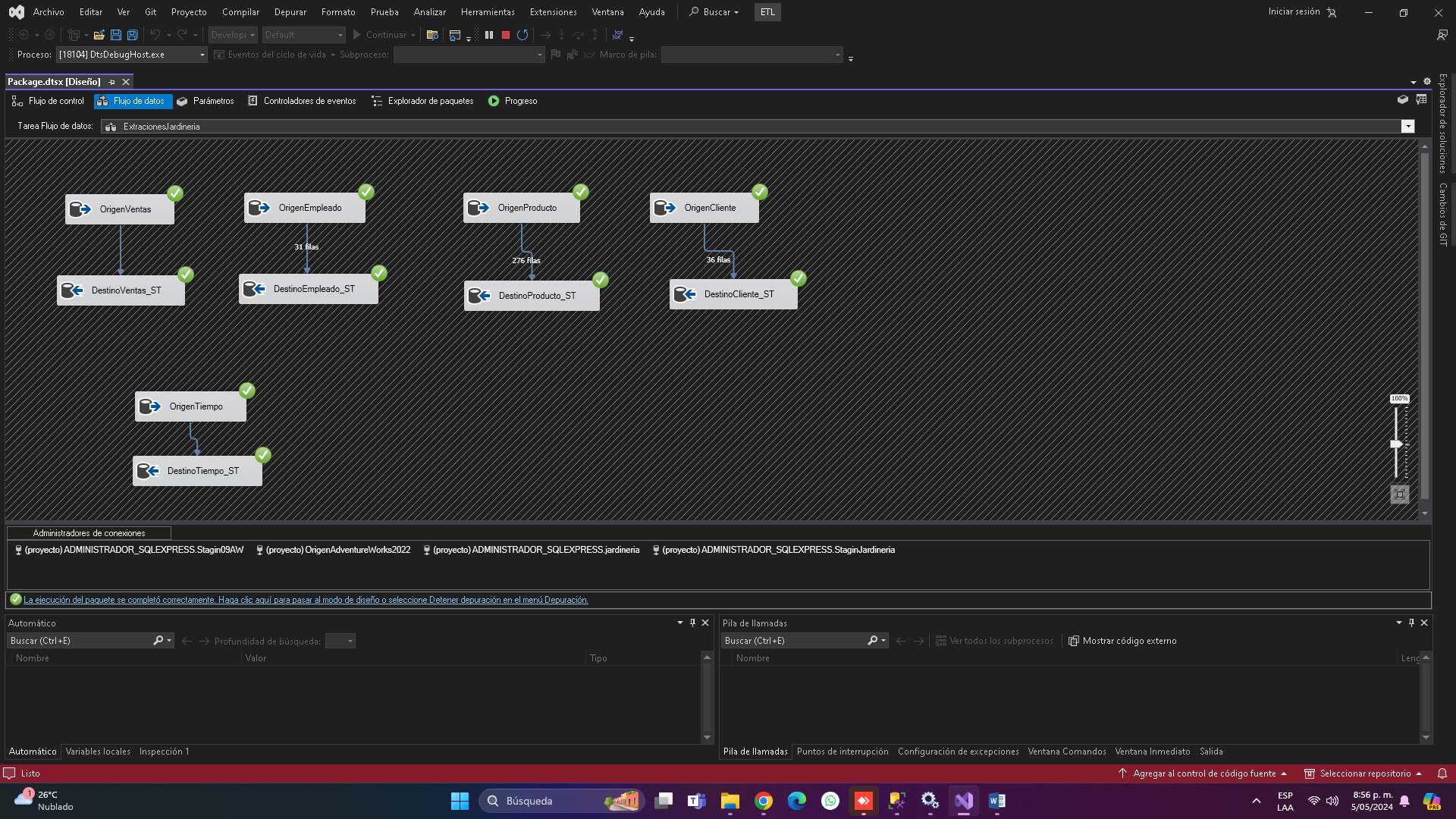
"puesto" nvarchar(50)

PRIMARY KEY( "id\_empleado")

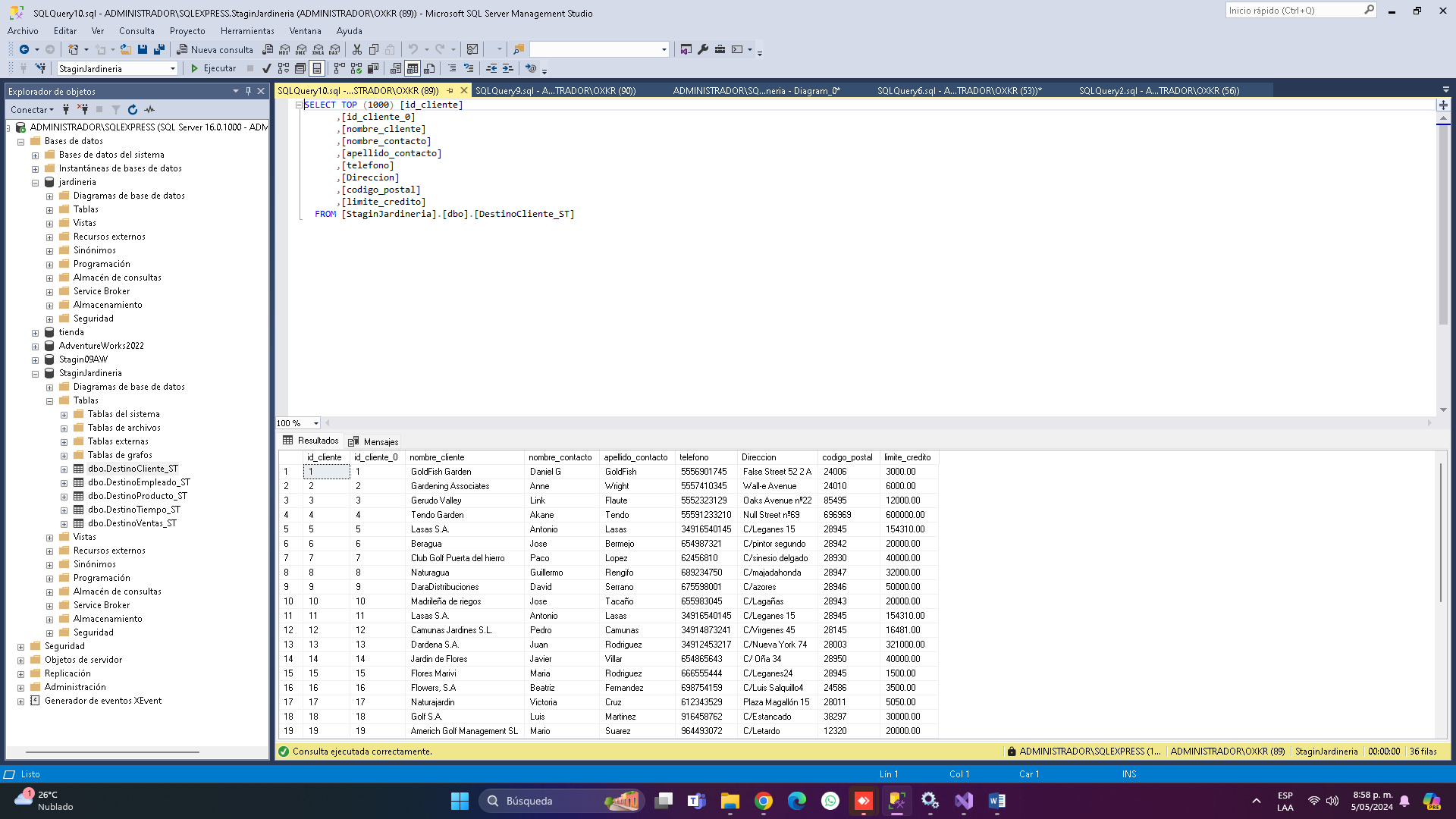
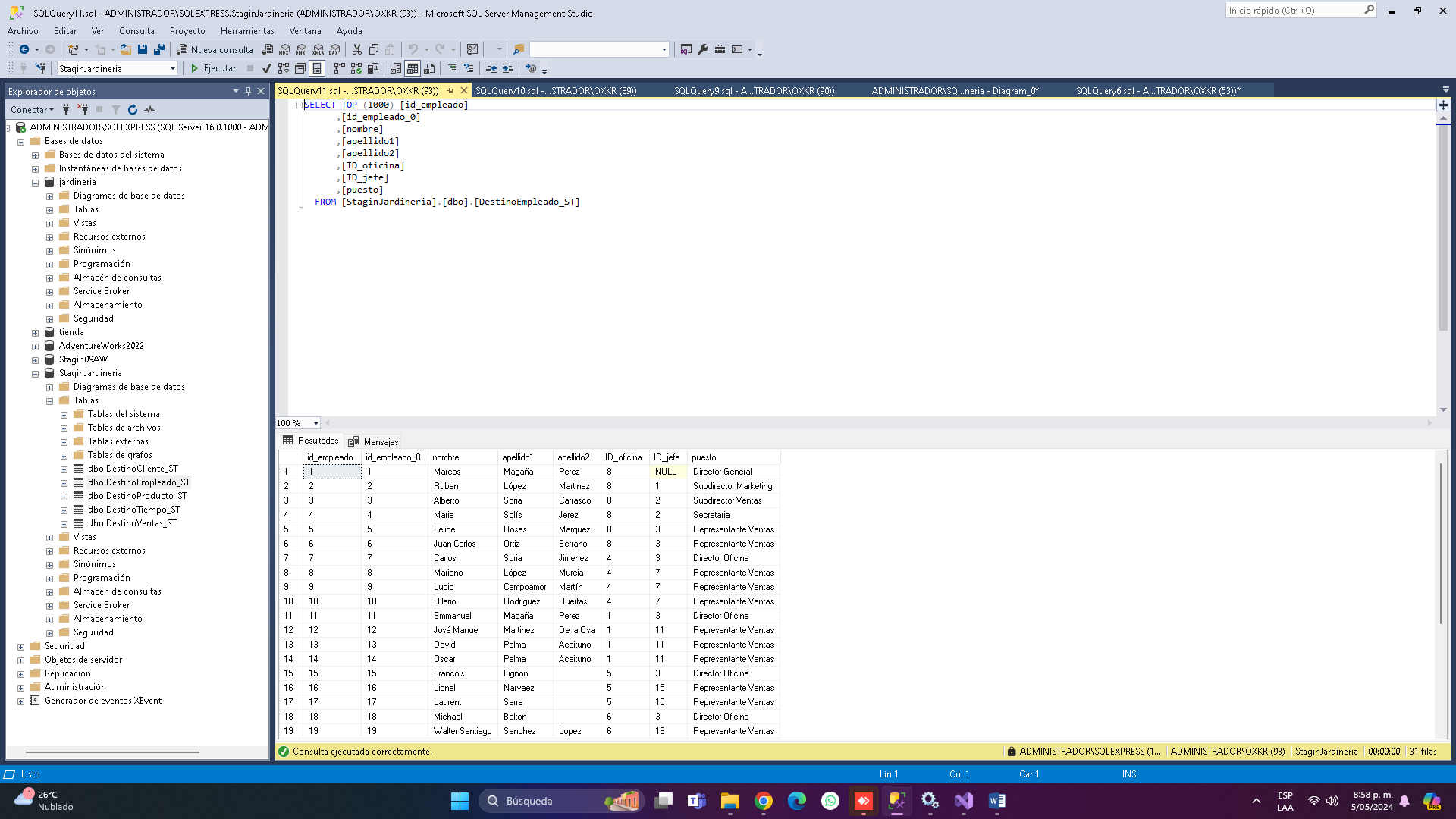
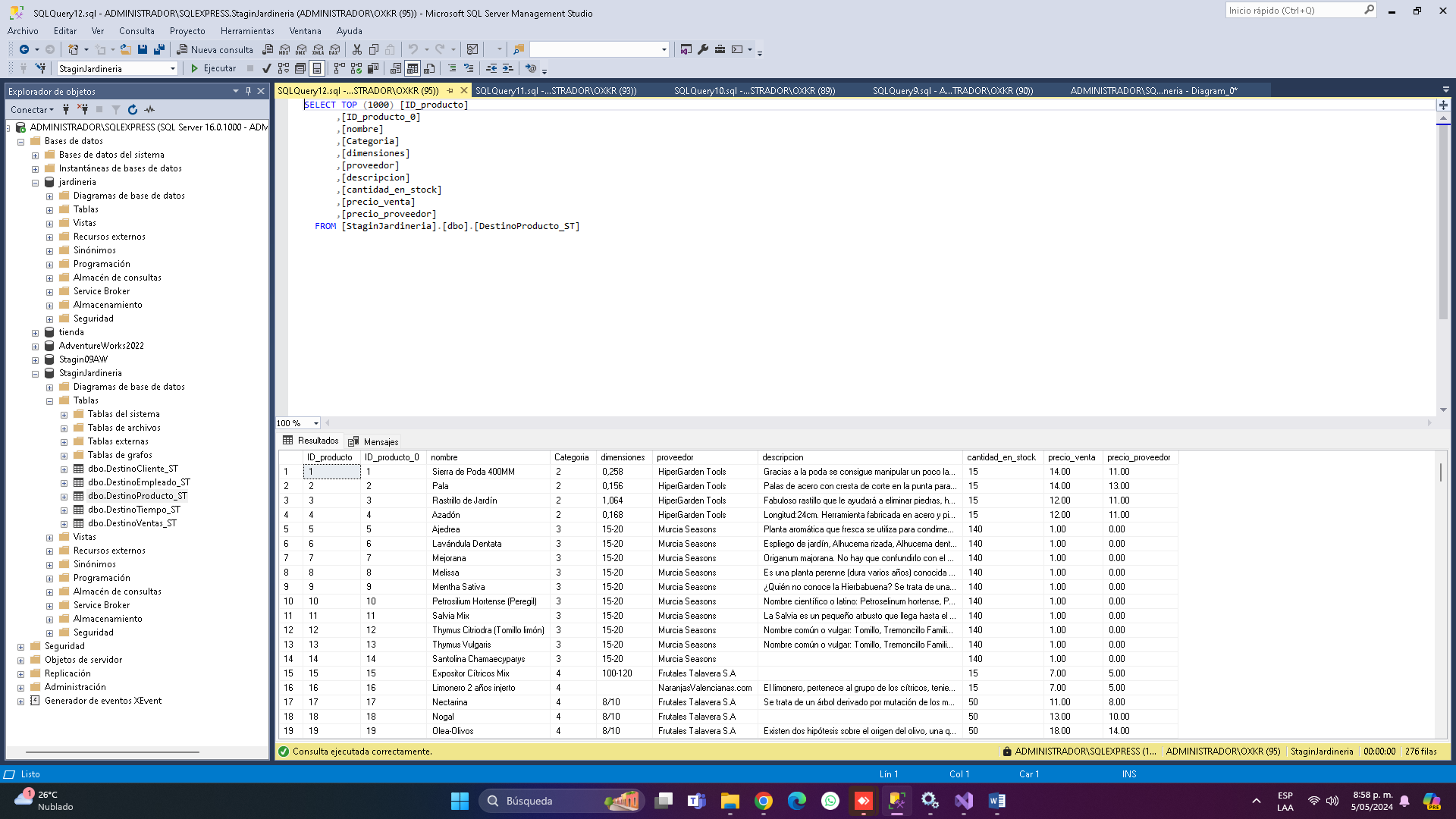
)



5.Ahora ejecutamos el Flujo de datos y vemos que todos corren correctamente, para luego mostrar en el sql server, con las consultas si las tablas quedaron bien en la base de StaginJardineria



6.Ejecutar las consultas y validar que los datos queden almacenados correctamente en la Base de datos *Staging*.

* + 
  + 
  + 

Como ven cada una de las tablas quedaron en la nueva base de datos StagingJardineria y la que tienen datos pasaron sus datos sin ningún problema

* + Construir el BK de ambas bases de datos.

En el siguiente drive dejaremos los links de las BK de las bases de datos y también el archivo sql de las consultas, además también estarán disponibles en los datos de la carga de la tarea cada uno por separado

<https://drive.google.com/drive/folders/1WGpQyeLamvntHOg1_smpRkS_jTj6OcEI?usp=sharing>

**Bibliografía**

* Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. John Wiley & Sons.
* Inmon, W. H., & Hackathorn, R. D. (1994). Using the Data Warehouse. John Wiley & Sons.
* Ralph Kimball et al., (2008). The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. John Wiley & Sons.
* García-Molina, H., Ullman, J. D., & Widom, J. (2008). Database Systems: The Complete Book. Pearson Education.