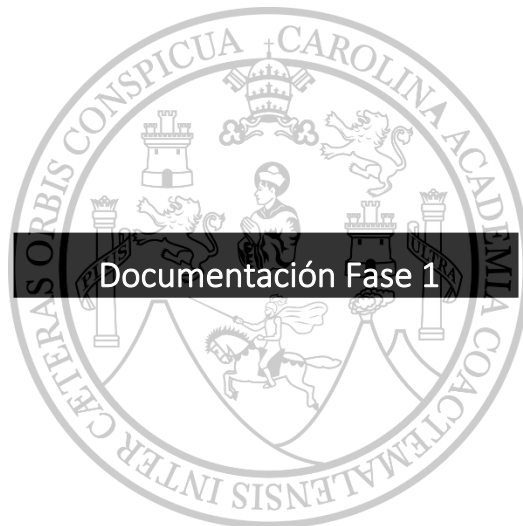


Universidad San Carlos de Guatemala

Ingeniería en Ciencias y Sistemas

Primer Semestre 2022

Seminario de Sistemas 2



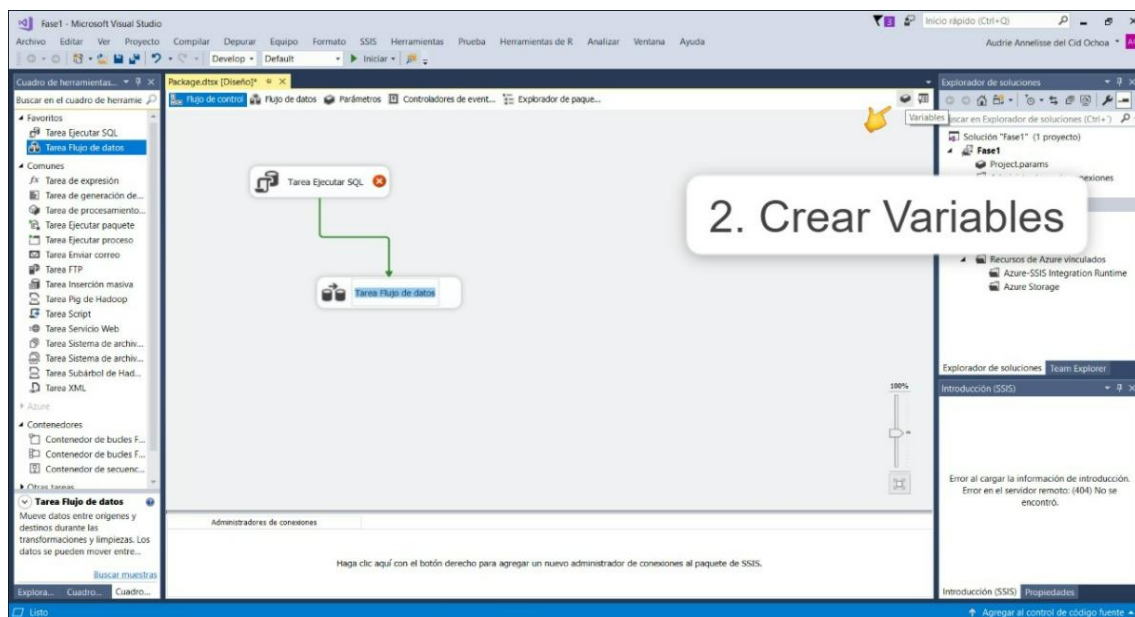
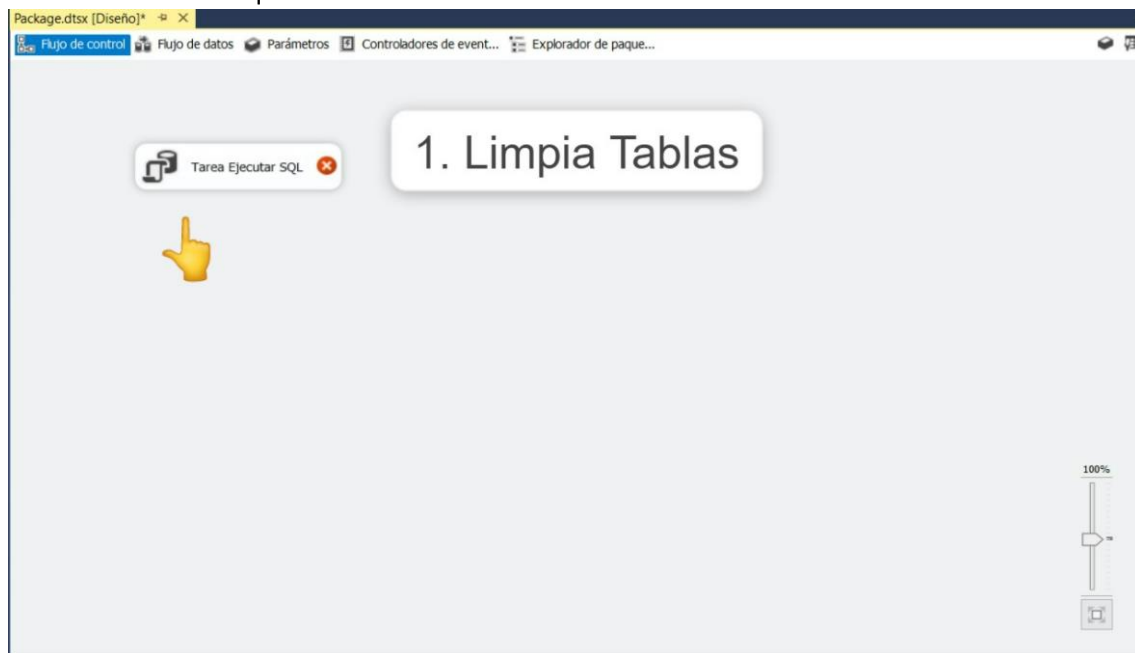
Audrie Annelisse del Cid Ochoa

201801263

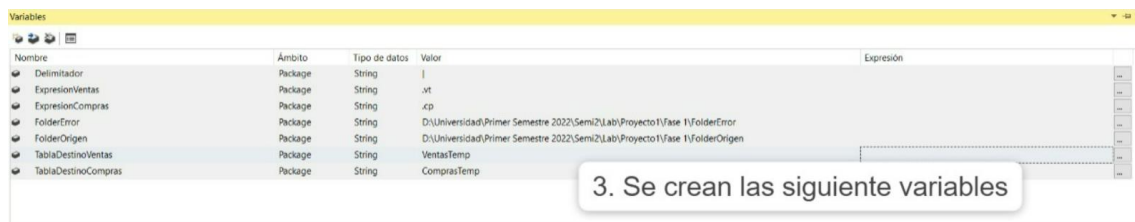
3012930900101

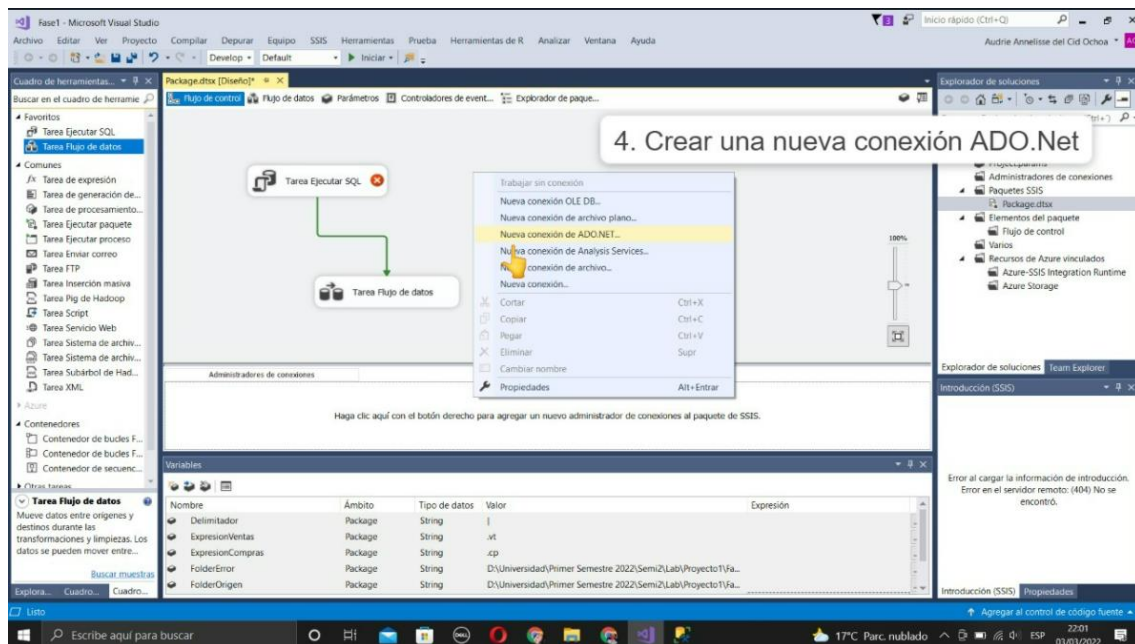
Flujo en SSIS

1. Colocar Execute sql task

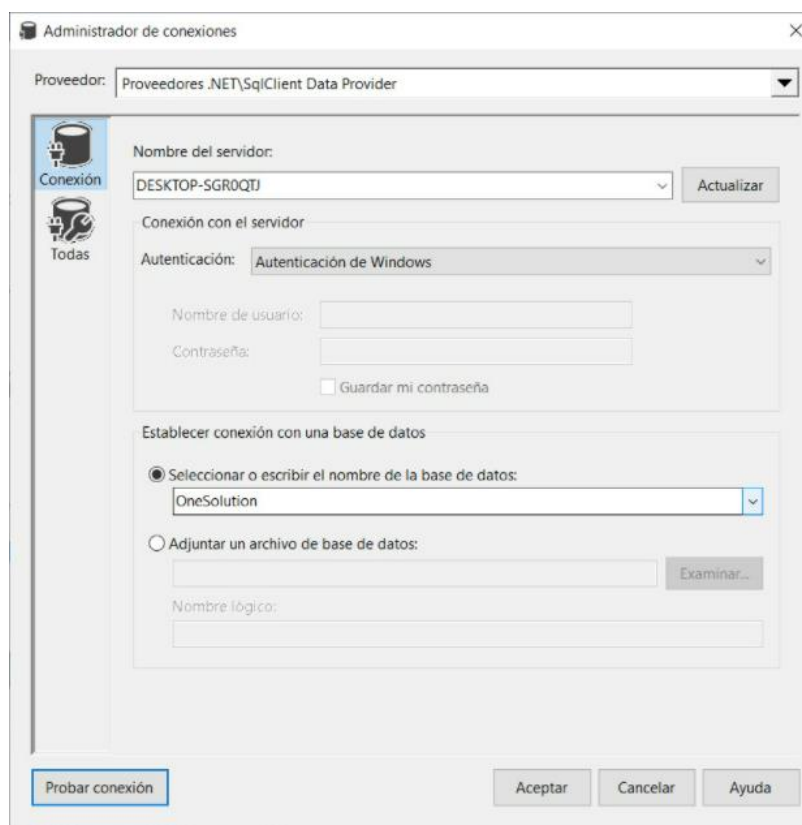


2. Las variables serán útiles al momento de mirar el archivo de entrada hacia la tabla temporal

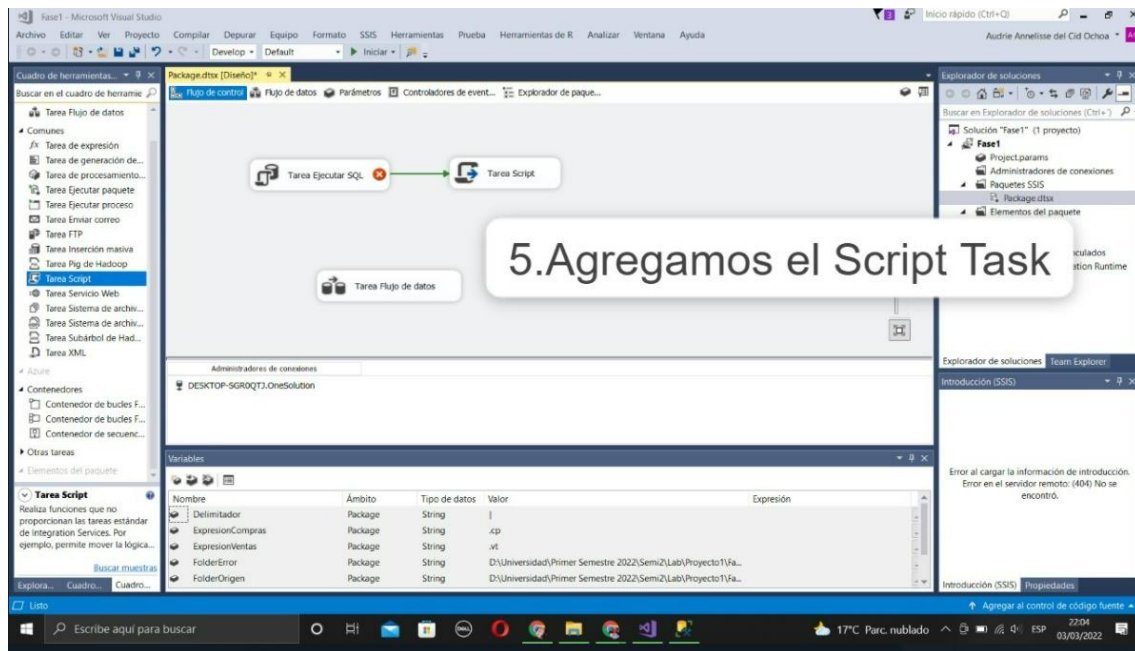




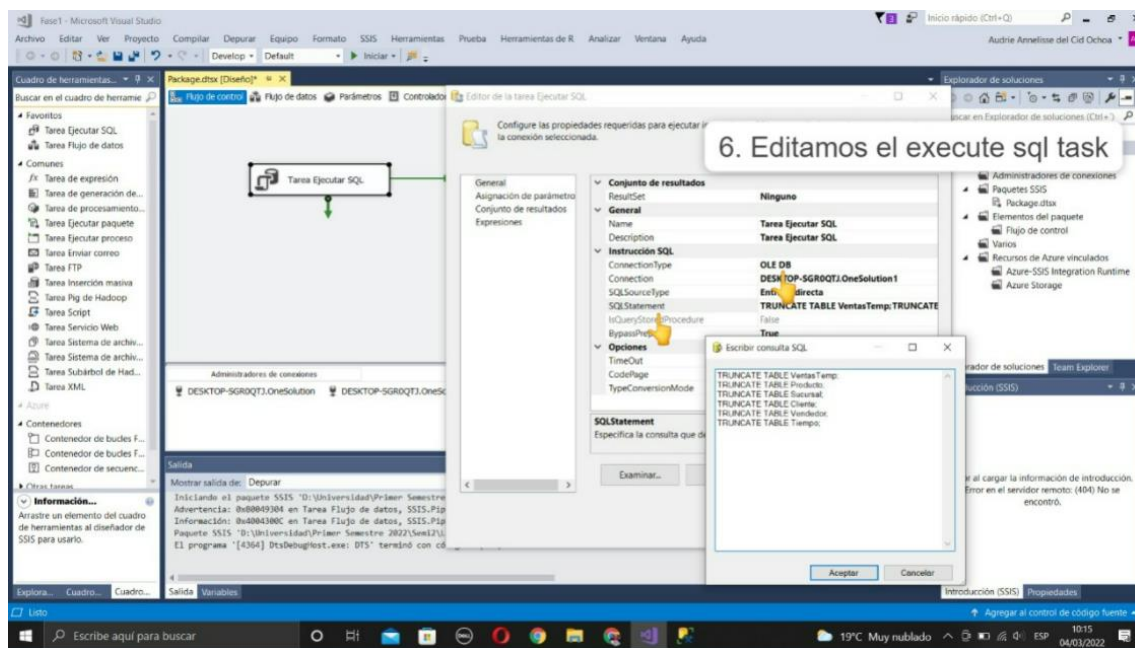
4. Así debe quedar la conexión



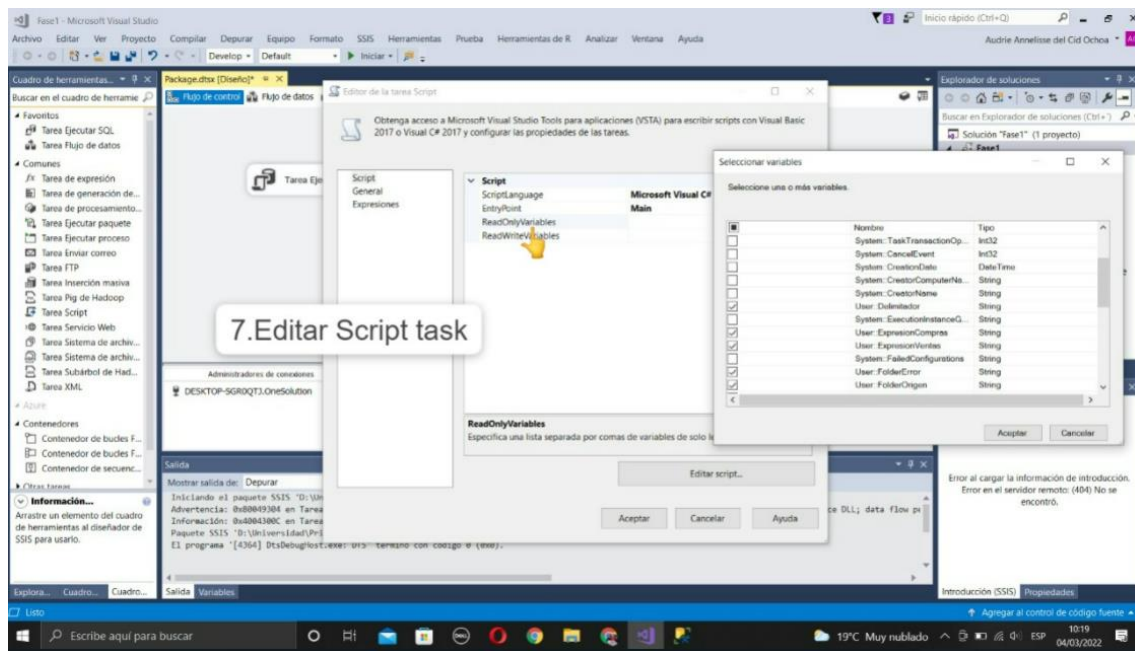
5. Este nos sirve para llenar los datos en nuestra tabla temporal



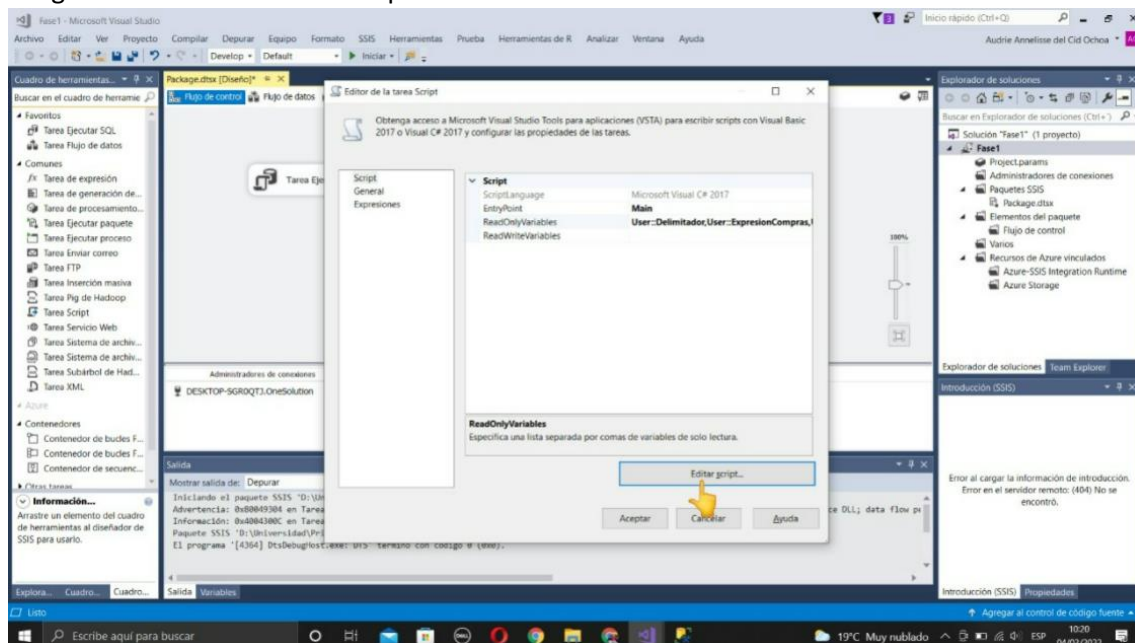
6. Cambiamos el tipo de conexión a OLE DB y escribimos los comandos sql que deseamos realizar. En este caso es de limpieza de tablas.



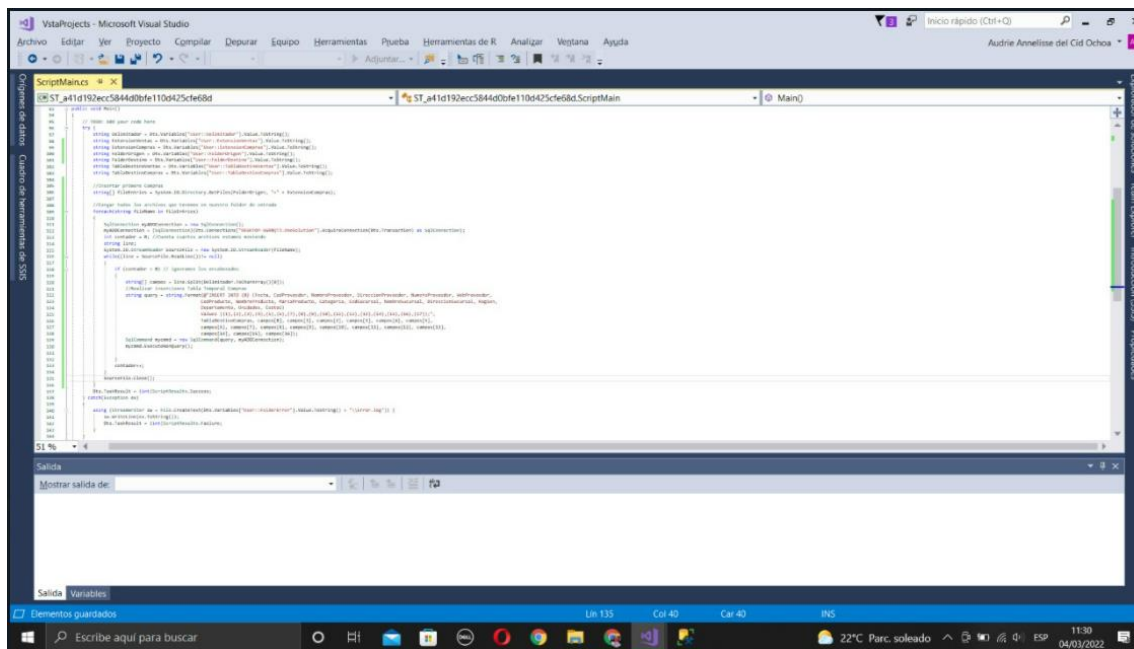
7. Editamos el script task y seleccionamos en ReadOnlyVariables, las variables que creamos



Luego damos click en editar script



Agregamos el Código en edit script. Nos servirá para llenar las tablas temporales



```
using System;
using System.Data;
using System.Data.SqlClient;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using Microsoft.SqlServer.Dts.Runtime;
using Microsoft.SqlServer.Dts.Tasks.ScriptTask;

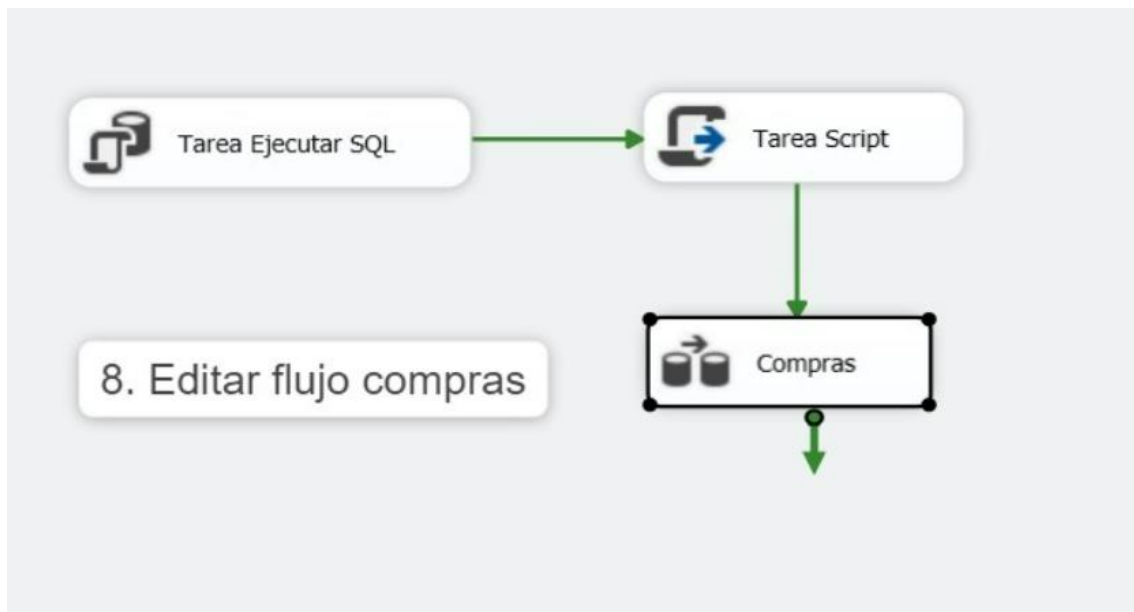
namespace Microsoft.SqlServer.Dts.Tasks.ScriptTask
{
    public class ScriptMain : ScriptTask
    {
        public override void Execute()
        {
            // Definición de variables
            string connectionString = "Data Source=.;Initial Catalog=tempdb;Integrated Security=True;User=root";
            string query = "SELECT * FROM tempdb.dbo.temp_table";

            // Conexión a la base de datos
            using (SqlConnection connection = new SqlConnection(connectionString))
            {
                connection.Open();

                // Ejecución de la consulta
                SqlCommand command = new SqlCommand(query, connection);
                SqlDataReader reader = command.ExecuteReader();

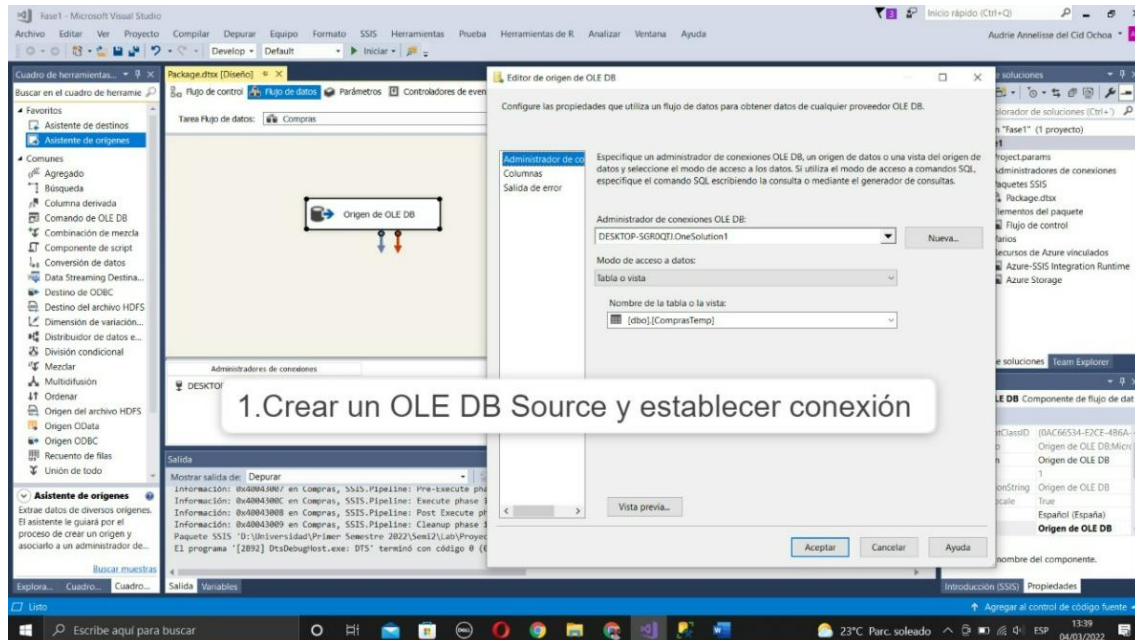
                // Procesamiento de los datos
                while (reader.Read())
                {
                    // Aquí se puede hacer algo con los datos de la consulta
                }
            }
        }
    }
}
```

8. Creamos un data flow task llamado Compras y procedemos a editar

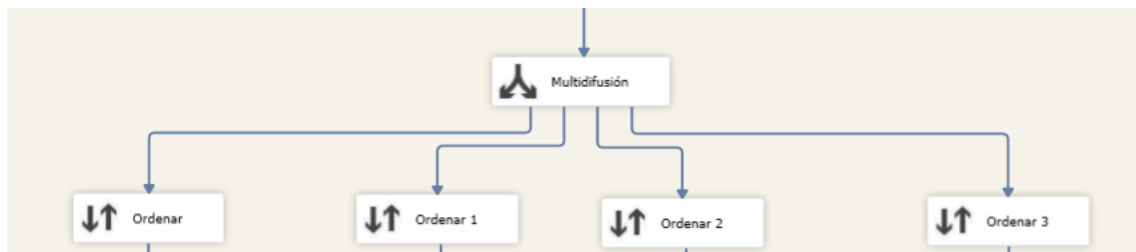


Proceso ETL de las Dimensiones de Compras

Creamos un OLE DB Source y establecemos conexión. En este caso se busca obtener los datos de la tabla Temporal Compras.



Luego dependiendo de la dimensión relacionada con los datos de compras que queremos llenar, su flujo irá cambiando, por lo tanto agregamos una herramienta de multidifusión que nos permitirá obtener los mismos datos por diferentes ramas creadas.



También se busca agregar la herramienta ordenar, la cual nos permitirá filtrar los datos que no están repetidos, guiándose por el código de cada dimensión.

A continuación se dará un breve resumen de lo realizado en cada rama de flujo.

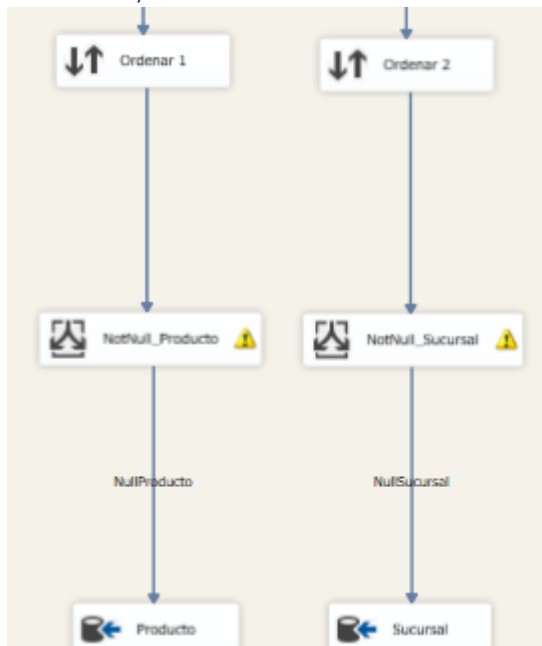
Proveedor



Después de ordenar utilizamos una herramienta par filtrar datos nulos, donde no se encuentre el código del proveedor. También mediante la conversión de string a datos numéricos comprobamos que el número telefónico del proveedor no tenga errores. También aplicamos un filtro para el proveedor web y comprobamos si sólo es 1 carácter y si dicho carácter es “S” o “N” y por último realizamos la conversión de datos, donde pasamos un varchar de 200 tanto para Numero proveedor y Web Proveedor, a sus correspondientes tipos de datos varchar[8] y varchar[1].

Una vez todos los datos han sido filtrados, se procede a cargar la dimensión Proveedor.

Producto y Sucursal



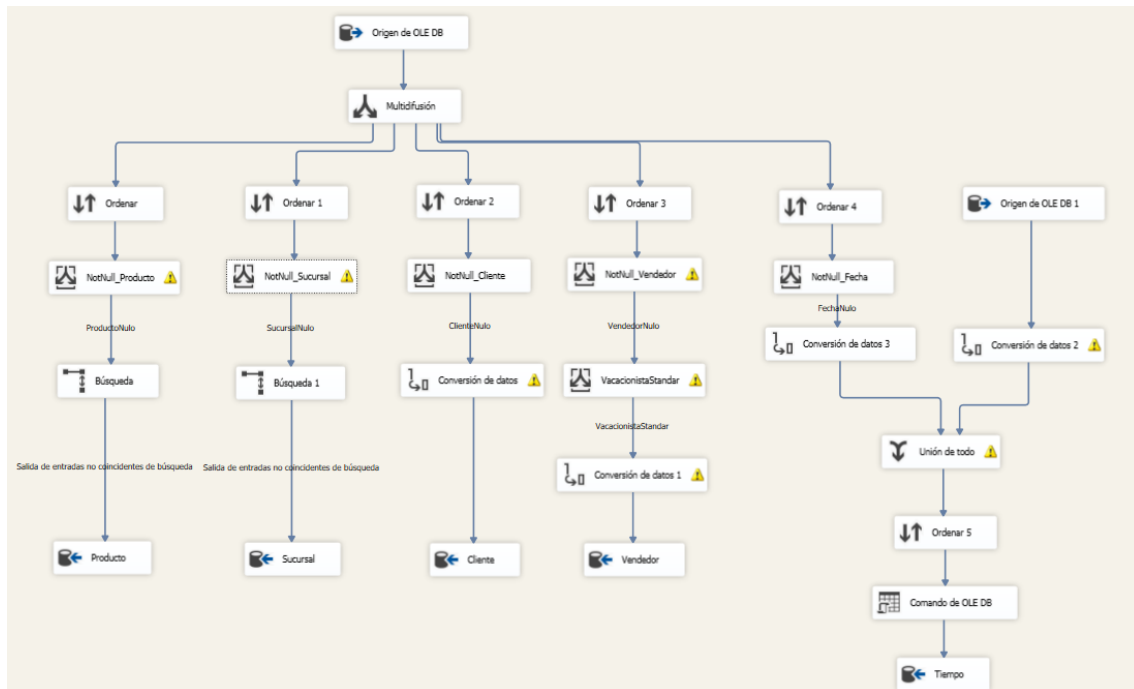
Una vez ordenados los datos, tanto para el flujo de Producto como para el de sucursal, se procede a verificar la existencia del código y filtrar en caso de que existan nulos. Una vez verificado esto se procede a cargar las dimensiones correspondientes..

Tiempo



Para este flujo una vez después de filtrar Fechas nulas, se procede a convertir el dato para su almacenamiento en la dimensión.

Proceso ETL de las dimensiones de Ventas

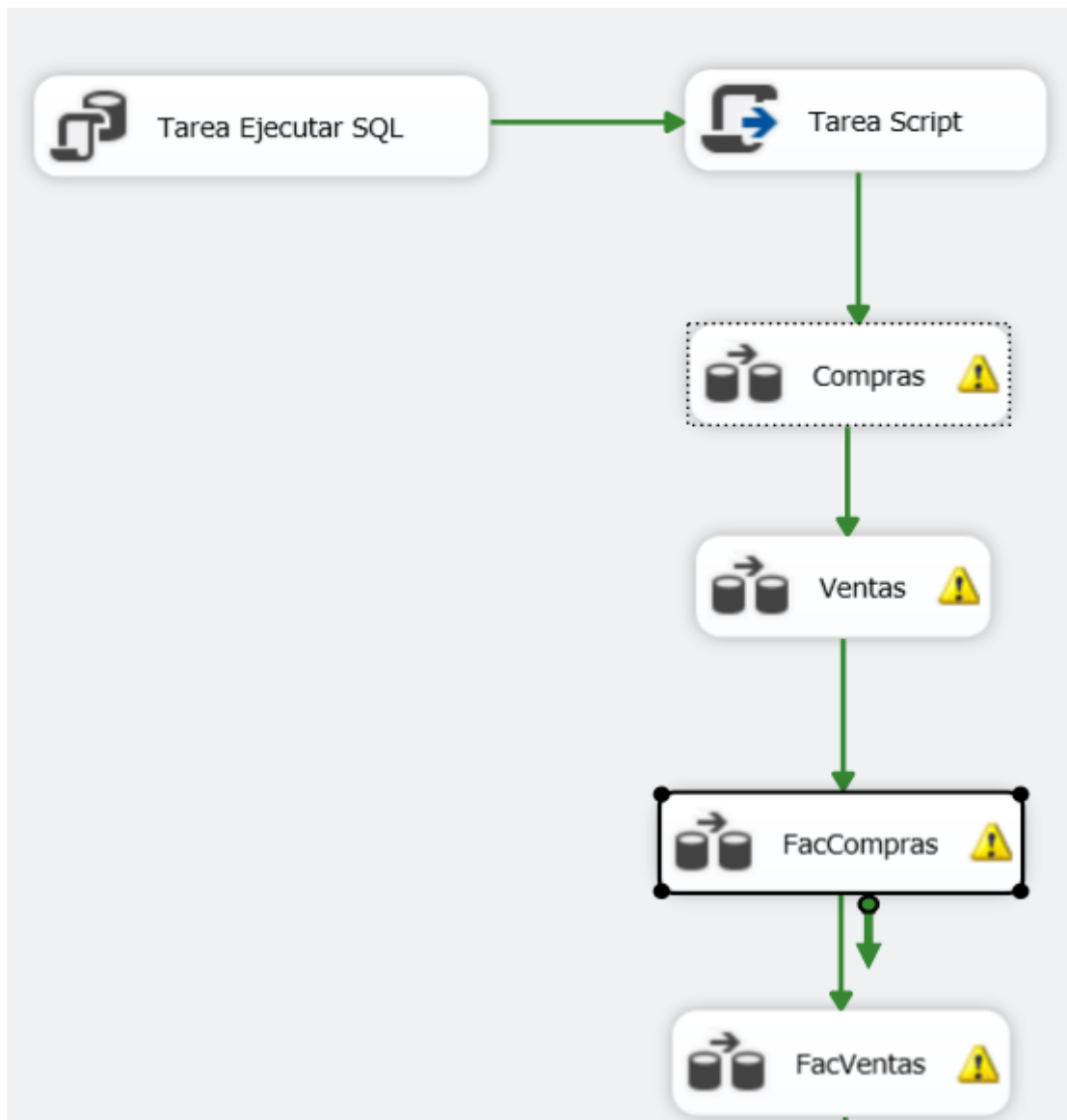


Este proceso es similar al detallado anteriormente, variando en algunas restricciones.

Lo que se puede destacar en este proceso es que para cada dimensión compartida como Producto, Sucursal y Tiempo, se necesita realizar una comparación, para verificar que los datos que desean ser almacenados no se dupliquen, por lo tanto se utiliza la herramienta de búsqueda, la cual permite realizar comparaciones con los datos insertados en la dimensión.

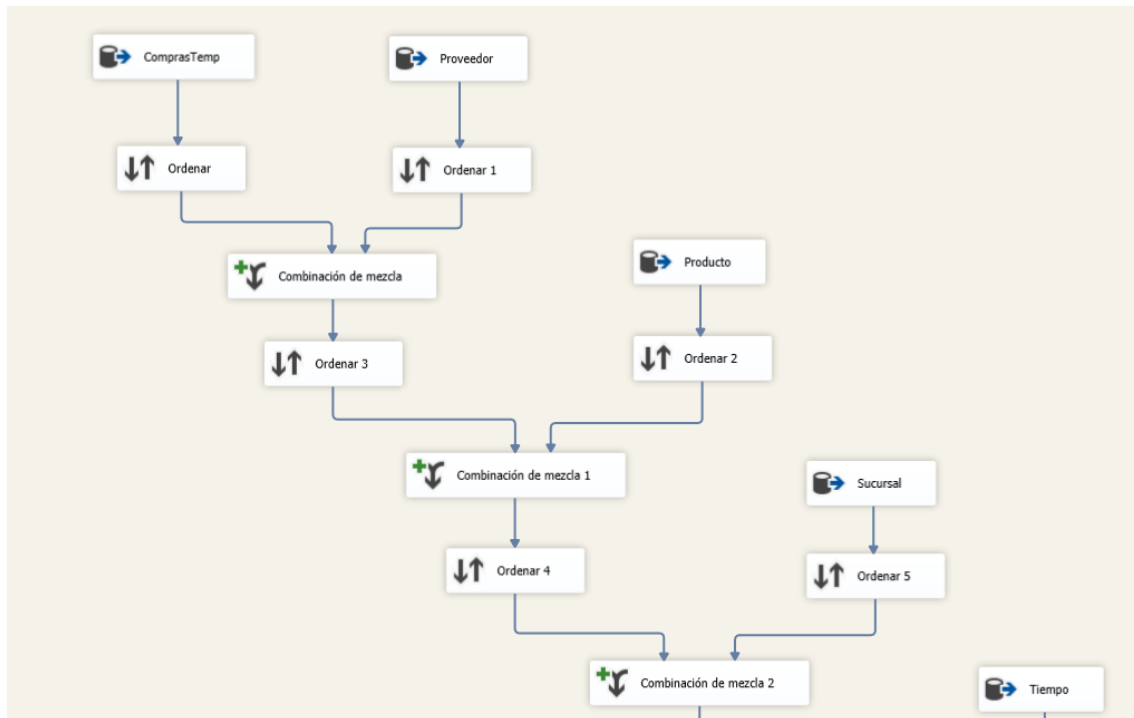
En este proceso también se obtuvo el llenado de datos para la dimensión de cliente y vendedor.

Una vez realizados los procesos ETL para carga de dimensiones, tanto para compras como ventas. Se procede a realizar el proceso ETL para la carga de las tablas de hechos existentes. En este caso son 2, la tabla de hechos de Compra y la tabla de Hechos de ventas.



Para ello se realizó lo siguiente.

ETL Fac Compras y Ventas

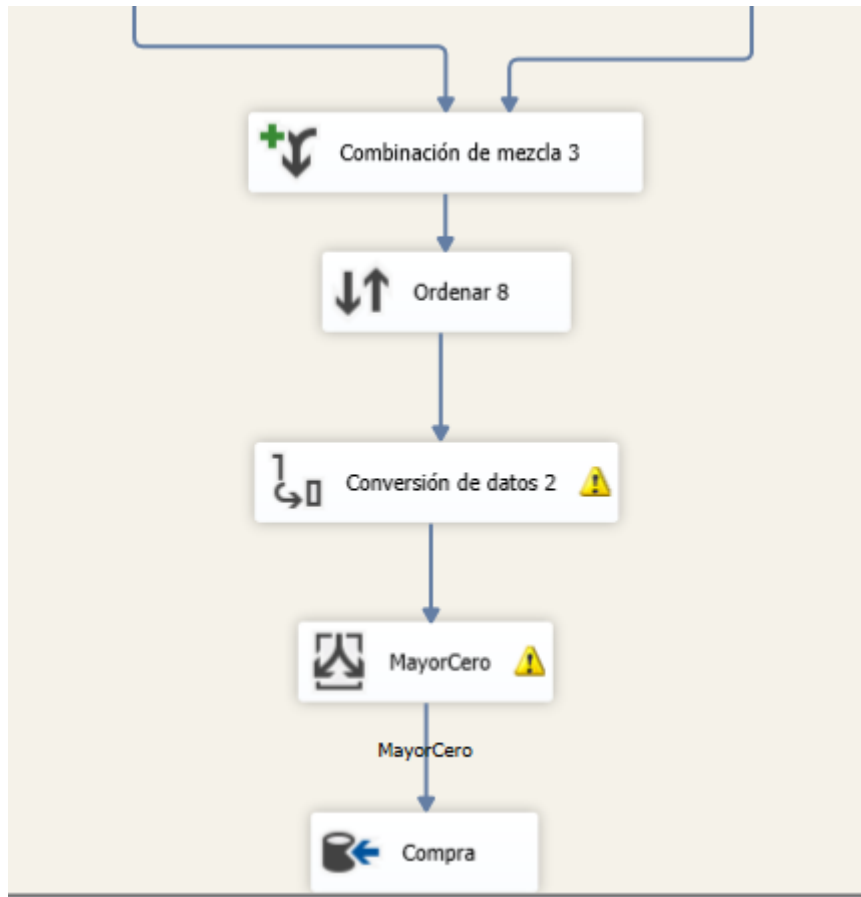


Tanto para la tabla de Hechos de compras como para la de ventas, se realizaron comparaciones entre los datos insertados y la tabla temporal que contiene todos los datos del archivo de entrada, esto para poder obtener los identificadores creados una vez insertados los datos en cada dimensión y así poder agregarlos a la tabla de hechos para crear la relación entre tabla de hechos y sus dimensiones.

Para esta búsqueda se utilizaron las siguientes herramientas:

- **Asistente de Orígenes:** Esta herramienta permitió obtener los datos de las tablas de dimensiones previamente llenadas.
- **Ordenar:** Esta herramienta sirvió tanto para filtrar los datos, así como para ordenar los datos como requisito de la herramienta combinación Mezcla.
- **Combinación Mezcla:** Esta herramienta permitió mezclar los datos, mediante un orden basados en un dato de las tablas que se desean unir.

Una vez obtenidos todos los datos necesarios para el llenado de la tabla, se procede a realizar las filtraciones y conversiones necesarias.



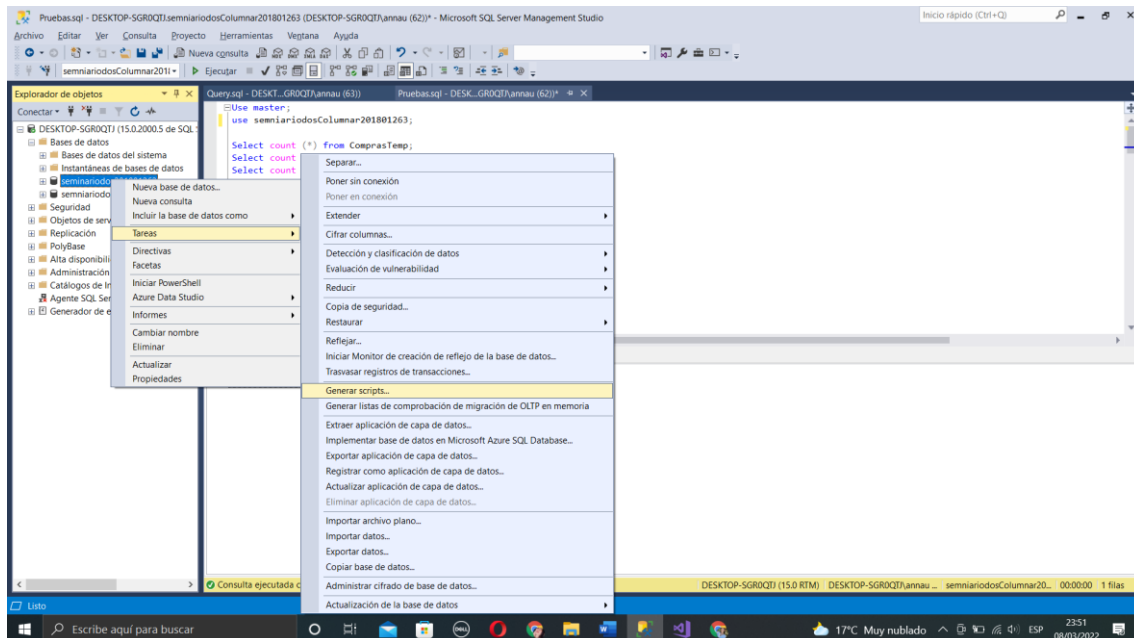
Tanto para la tabla de hechos de compras como ventas se realizaron las filtraciones para los datos numéricos, como la conversión de datos de cadenas a datos numéricos y la comprobación de datos duplicados. Luego se procedió a llenar cada tabla de hecho.

Una vez cargados todas las tablas de hechos, se termina el flujo y se procede a la clonación de la base de Datos y la conversión de tablas almacenadas mediante almacenamiento por fila a almacenamiento columnar.

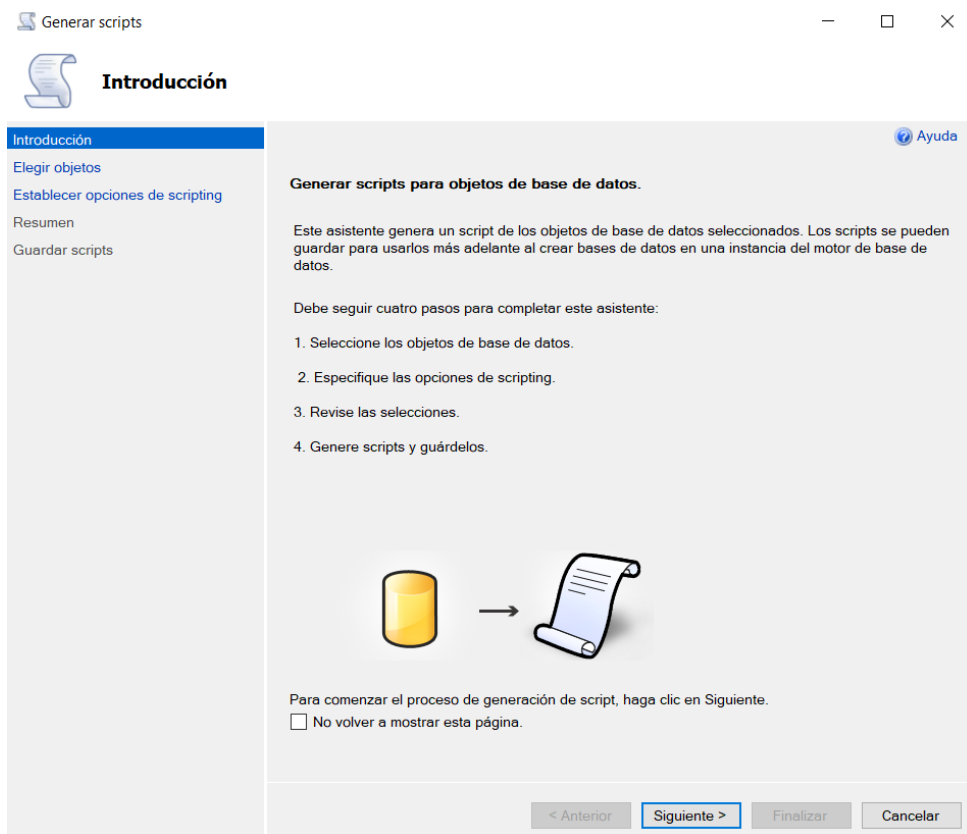
Clonación Base de Datos

Para la clonación de la base de datos, se realizaron los siguientes pasos.

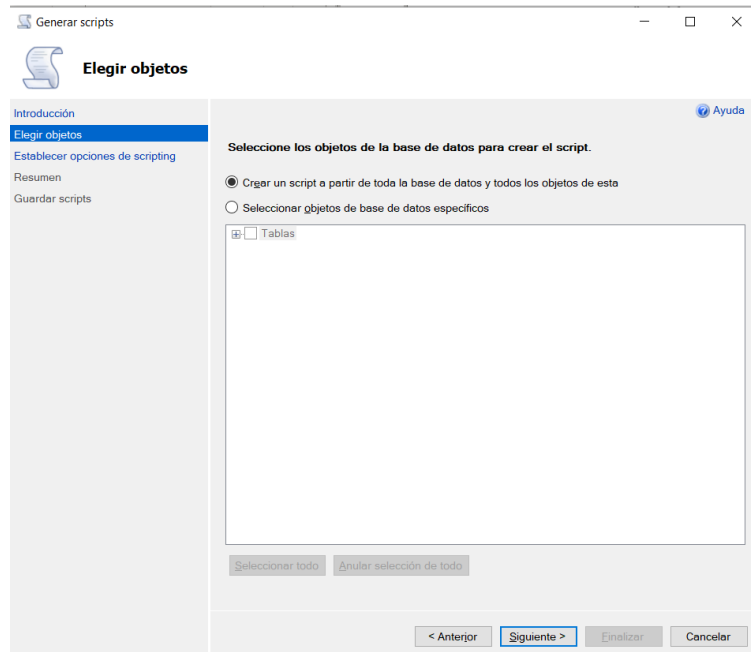
1. Seleccionamos la base de datos que deseamos, elegimos la opción de tareas y la opción Generar scripts



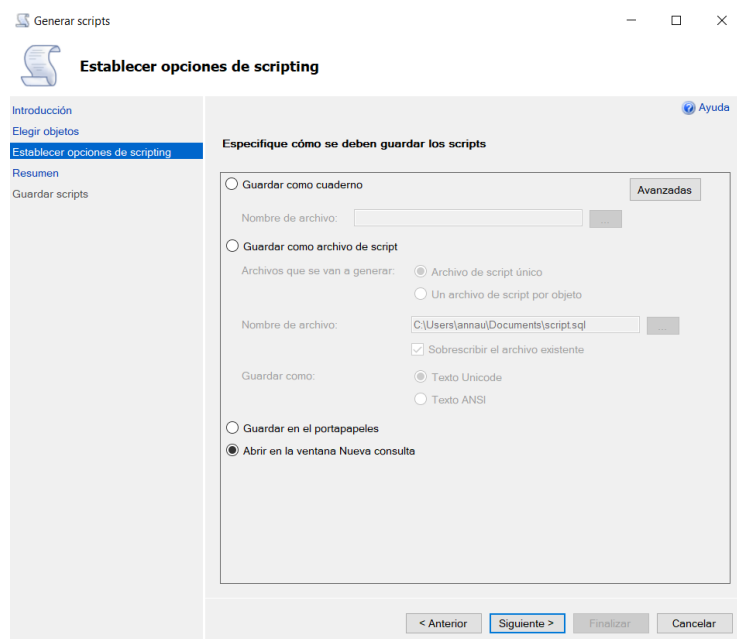
Nos aparecerá el siguiente asistente



2. Damos click en siguiente y nos aparecerá lo siguiente. Elegimos que deseamos generar en el script



Elegimos como deseamos mostrar o guardar el script



Una vez terminado esto, podemos cargar el script a nuestra nueva base de datos.

Creación de índices Columnares

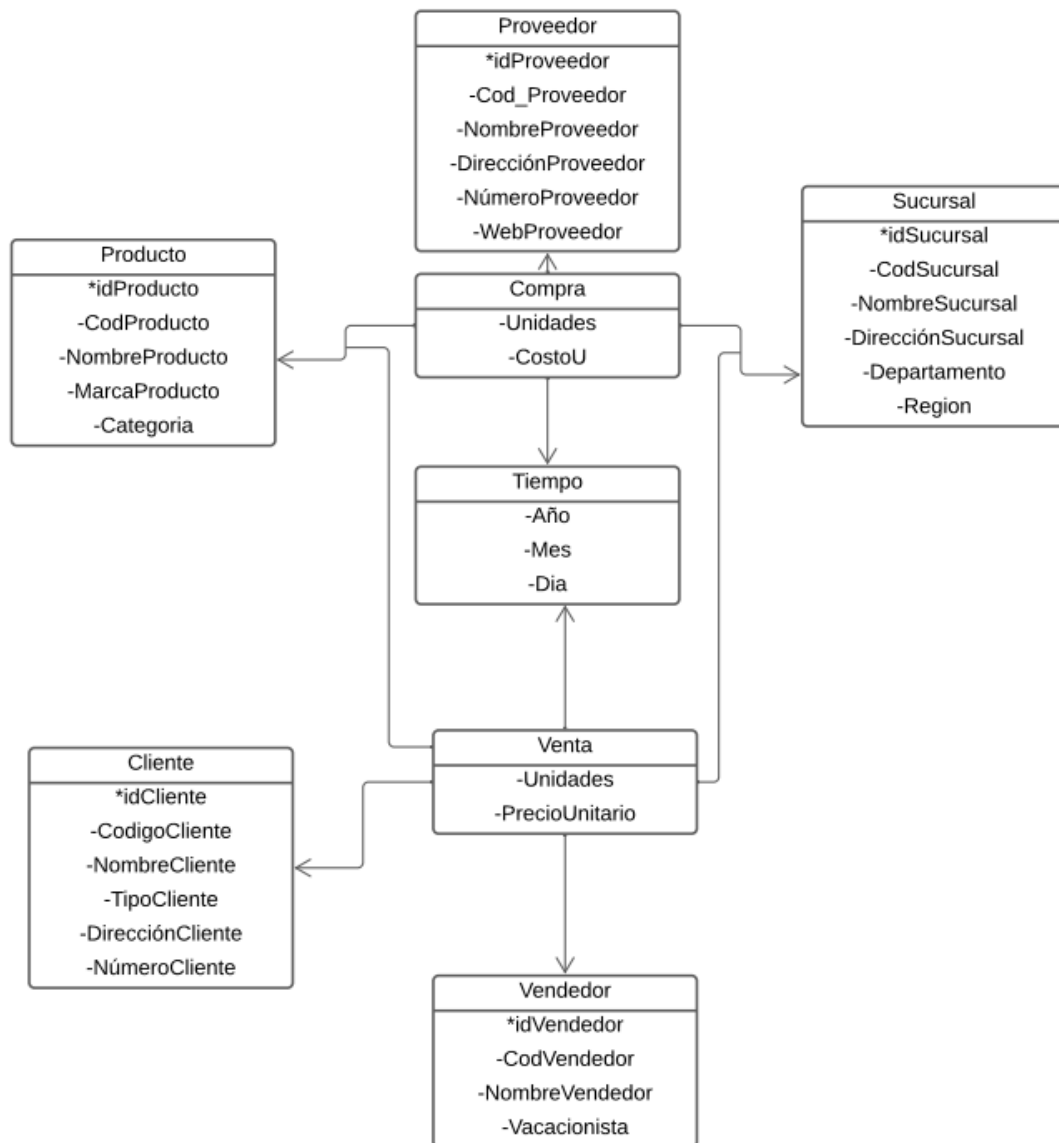
Para finalizar, una vez clonada nuestra base de datos procedemos a crear los índices para el almacenamiento columnar en nuestra nueva base de datos.

Para esto se utilizó el siguiente script.

```
use seminariosColumnar201801263;

CREATE NONCLUSTERED COLUMNSTORE INDEX ProveedorC ON dbo.Proveedor (idProveedor,CodProveedor, NombreProveedor, DireccionProveedor, NumeroProveedor, WebProveedor );
CREATE NONCLUSTERED COLUMNSTORE INDEX ProductoC ON dbo.Producto (idProducto, CodProducto, NombreProducto, MarcaProducto, Categoria);
CREATE NONCLUSTERED COLUMNSTORE INDEX ClienteC ON dbo.Cliente (idCliente,CodigoCliente, NombreCliente, TipoCliente, DireccionCliente, NumeroCliente);
CREATE NONCLUSTERED COLUMNSTORE INDEX VendedorC ON dbo.Vendedor(idVendedor, CodVendedor, NombreVendedor, Vacacionista);
CREATE NONCLUSTERED COLUMNSTORE INDEX SucursalC ON dbo.Sucursal(idSucursal, CodSucursal, NombreSucursal, DireccionSucursal,Departamento,Region);
CREATE NONCLUSTERED COLUMNSTORE INDEX TiempoC ON dbo.Tiempo (idTiempo, Fecha);
CREATE NONCLUSTERED COLUMNSTORE INDEX CompraC ON dbo.Compra (idCompra, idProveedor, idProducto, idTiempo, idSucursal, Unidades, CostoU);
CREATE NONCLUSTERED COLUMNSTORE INDEX VentaC ON dbo.Venta (idVenta, idProducto, idCliente, idVendedor, idSucursal, idTiempo,Unidades, PrecioUnitario);
```

Modelo Implementado



Se realizaron las siguientes tablas:

- **Tabla de Hechos:** Compra y Venta
- **Tabla Dimensión:** Proveedor, Producto, Sucursal, Cliente, Vendedor y Tiempo

Dado que se consideró necesaria la creación de dos tablas de hechos, podemos tomar esta característica para identificar este modelo como un modelo Constelación. Además, que dichas tablas de hechos también comparten tablas dimensionales como lo podemos observar con las dimensiones de Producto, Sucursal y Tiempo.