

**Universidad de San Carlos de Guatemala**  
**Facultad de Ingeniería**  
**Laboratorio Seminario de Sistemas 2 sección A**  
**AUX. Jose Fernando Alvarez Morales**



## **PROYECTO 1 FASE 1 SG-FOOD**

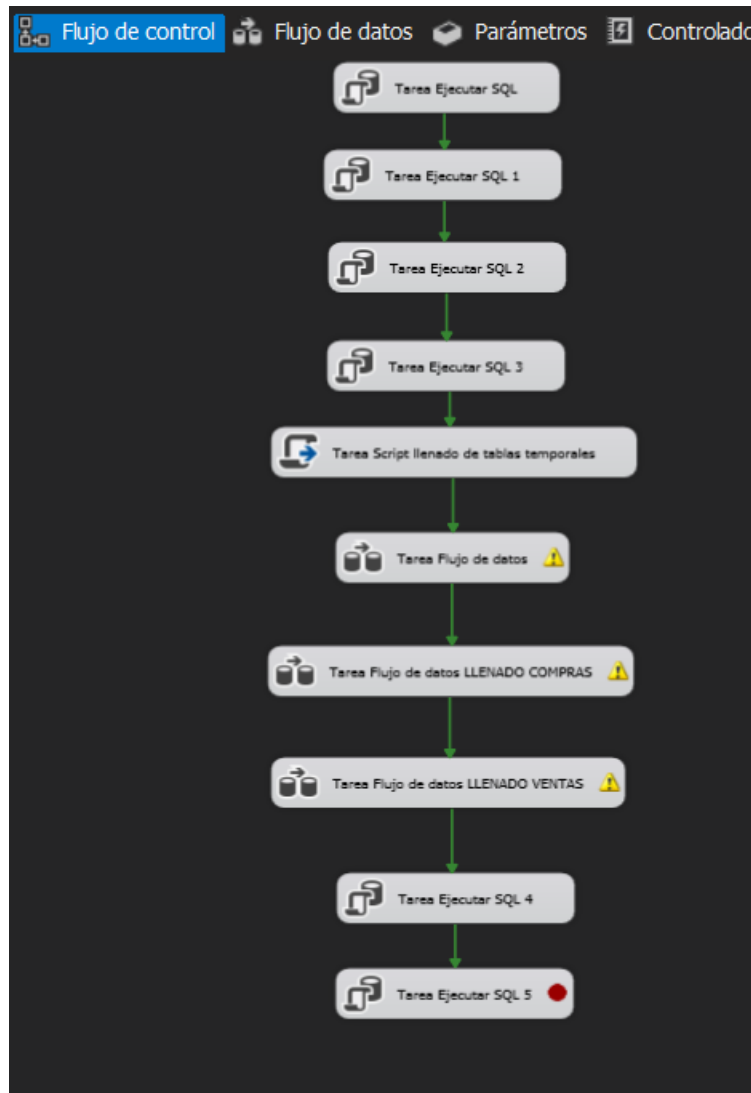
**Luis Angel Barrera Velásquez**

**RA: 202010223**

Guatemala, 11 de marzo de 2024

## Explicación de la solución ETL

Comenzando con el Flujo de control del proyecto el cual consta de diferentes herramientas SSIS la cual las primeras 3 de “Tarea Ejecutar SQL” son para reiniciar las tablas temporales o pivote de las diferentes fuentes de base de datos y la cuarta es para eliminar todas las tablas del modelo datawarehouse, cabe mencionar que se tienen 3 base de datos una en MYSQL llamada fuente1 otra en SQL server llamada fuente2 y la ultima también en SQL server que es la que contiene el modelo datawarehouse llamada seminario2\_202010223.



También se tiene la herramienta “Tarea Script” la cual es la extracción de los datos de los archivos .comp y .vent y son cargados a las bases de datos de fuente1 y fuente2 a su respectiva tablas temporales. Las tablas que se crearon son las siguientes:

## MYSQL:

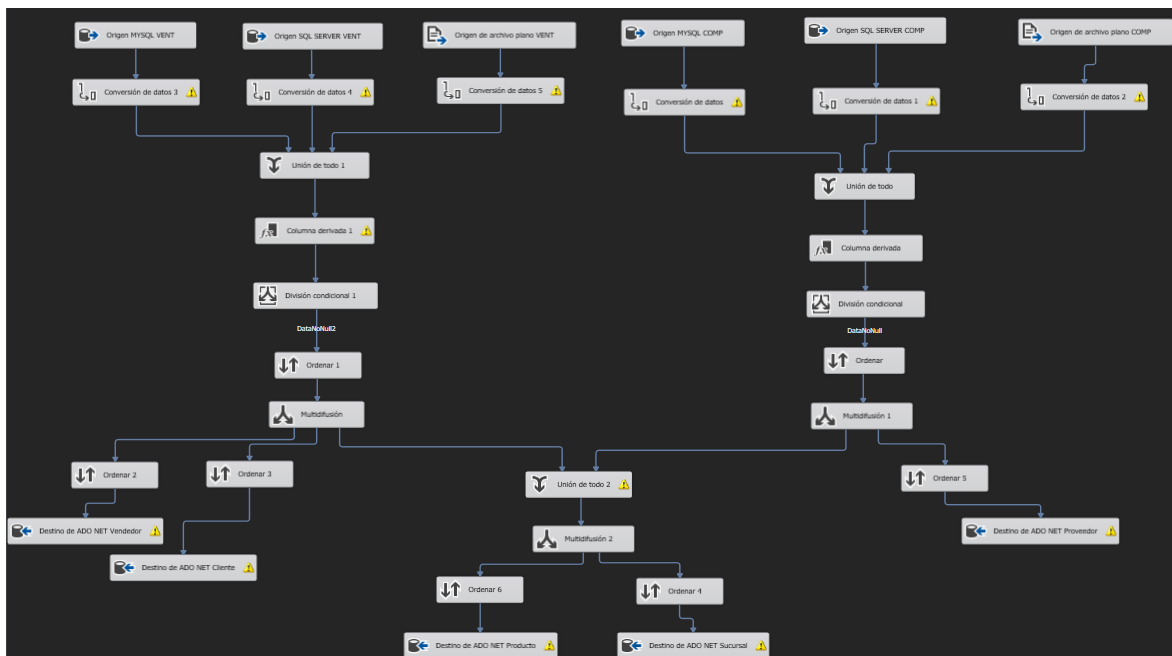
tempcomp
Fecha VARCHAR(255)
CodProveedor VARCHAR(255)
NombreProveedor VARCHAR(255)
DireccionProveedor VARCHAR(255)
NumeroProveedor VARCHAR(255)
WebProveedor VARCHAR(255)
CodProducto VARCHAR(255)
NombreProducto VARCHAR(255)
MarcaProducto VARCHAR(255)
Categoria VARCHAR(255)
SodSuSursal VARCHAR(255)
NombreSucursal VARCHAR(255)
DireccionSucursal VARCHAR(255)
Region VARCHAR(255)
Departamento VARCHAR(255)
Unidades VARCHAR(255)
CostoU VARCHAR(255)

tempvent
Fecha VARCHAR(255)
CodigoCliente VARCHAR(255)
NombreCliente VARCHAR(255)
TipoCliente VARCHAR(255)
DireccionCliente VARCHAR(255)
NumeroCliente VARCHAR(255)
CodVendedor VARCHAR(255)
NombreVendedor VARCHAR(255)
Vacacionista VARCHAR(255)
CodProducto VARCHAR(255)
NombreProducto VARCHAR(255)
MarcaProducto VARCHAR(255)
Categoria VARCHAR(255)
SodSuSursal VARCHAR(255)
NombreSucursal VARCHAR(255)
DireccionSucursal VARCHAR(255)
Region VARCHAR(255)
Departamento VARCHAR(255)
Unidades VARCHAR(255)
PrecioUnitario VARCHAR(255)

## SQLSERVER:

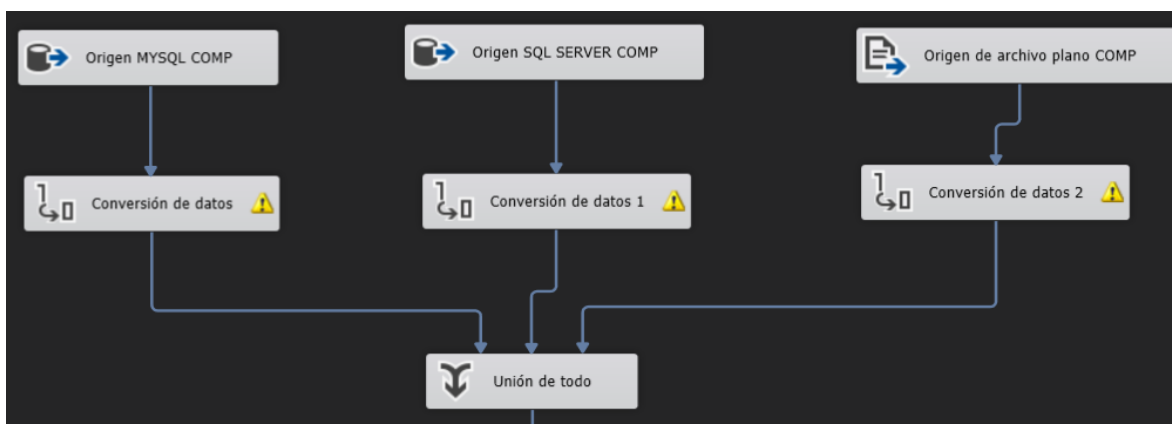
tempComp	tempVent
ABC Fecha	ABC Fecha
ABC CodProveedor	ABC CodigoCliente
ABC NombreProveedor	ABC NombreCliente
ABC DireccionProveedor	ABC TipoCliente
ABC NumeroProveedor	ABC DireccionCliente
ABC WebProveedor	ABC NumeroCliente
ABC CodProducto	ABC CodVendedor
ABC NombreProducto	ABC NombreVendedor
ABC MarcaProducto	ABC Vacacionista
ABC Categoria	ABC CodProducto
ABC SodSuSursal	ABC NombreProducto
ABC NombreSucursal	ABC MarcaProducto
ABC DireccionSucursal	ABC Categoria
ABC Region	ABC SodSuSursal
ABC Departamento	ABC NombreSucursal
ABC Unidades	ABC DireccionSucursal
ABC CostoU	ABC Region
	ABC Departamento
	ABC Unidades
	ABC PrecioUnitario

Después tenemos un Flujo de datos donde lo que hacemos es la transformación de los datos. A continuación, se muestra el flujo:

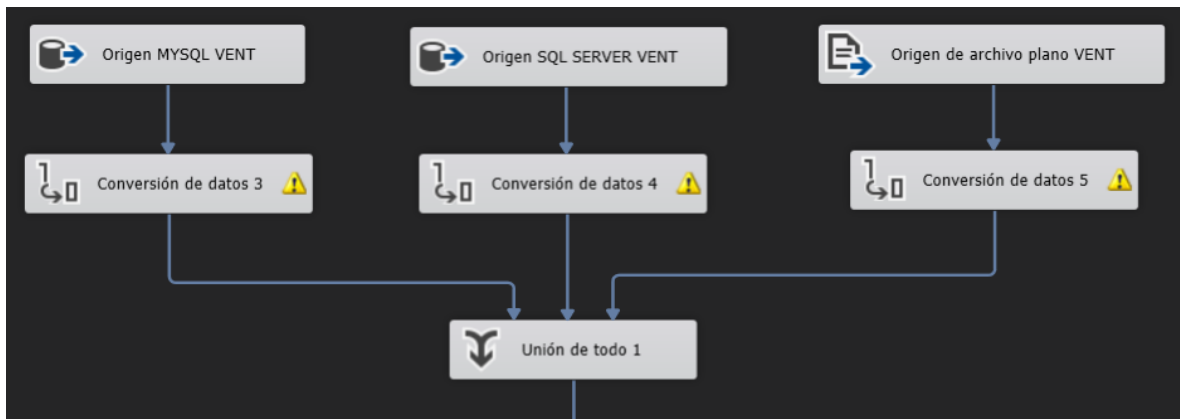


Se divide en partes por ejemplo en la parte superior se hace la extracción de las tablas temporales de las bases de datos fuente1 y fuente2 y el archivo de texto plano de los dos tipos de archivos, luego se hace la conversión de datos y la unión de los mismos:

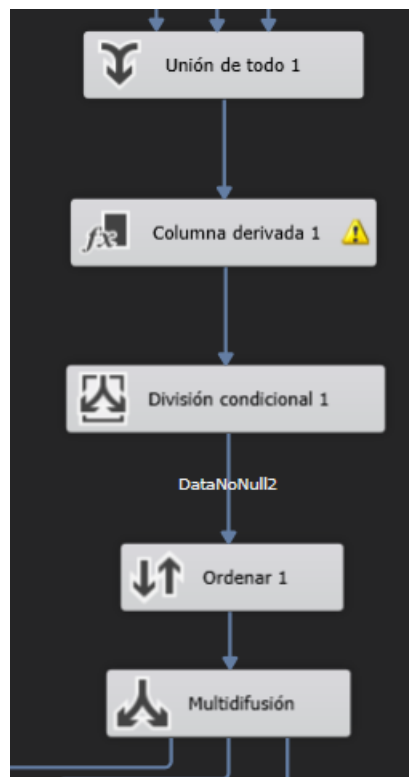
Fuentes de compra:



Fuentes de venta:



Luego procedemos a limpiar los datos que es parte de la transformación la cual se hace con el proceso derivar columnas y la división condicional, por ultimo ya teniendo limpio los datos se realiza una eliminación de duplicados usando la herramienta “Ordenar”



Por ultimo después de la transformación viene la carga, la cual se llenan las tablas de dimensión:

Tablas dimensionales de Vendedor y Cliente:

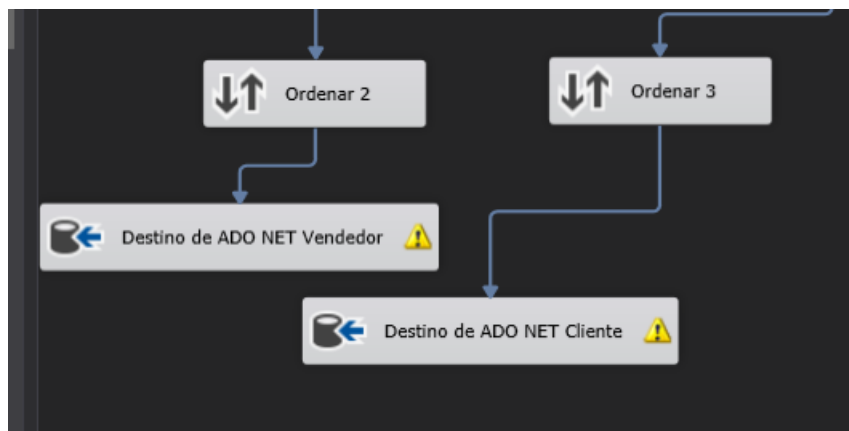
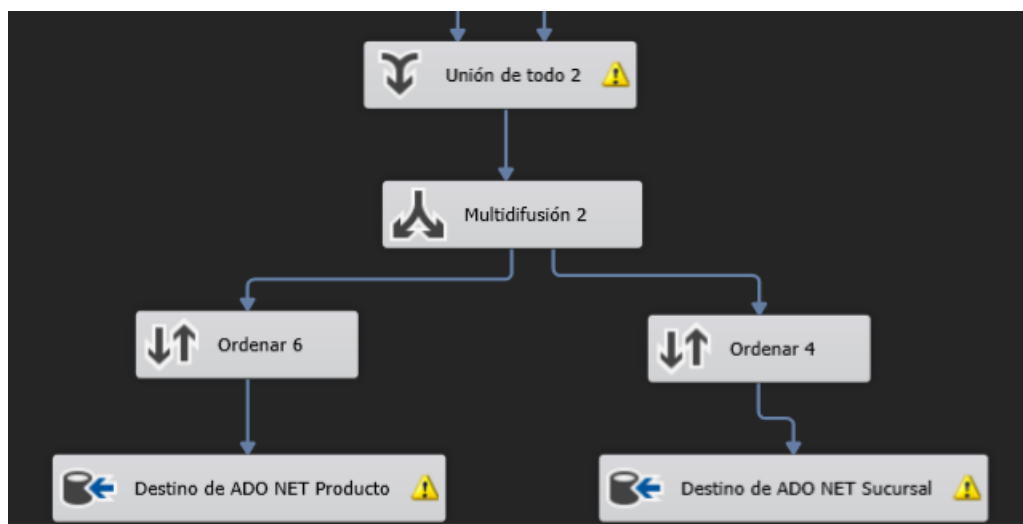


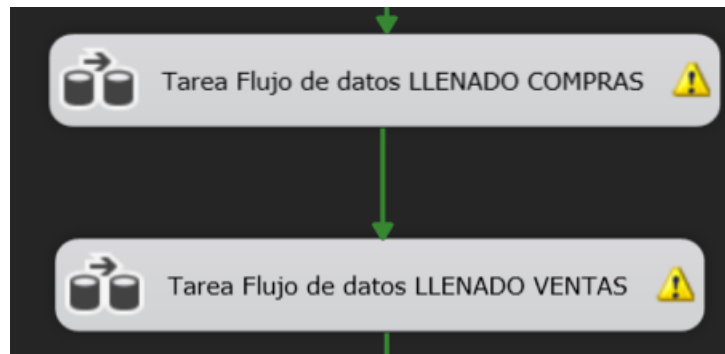
Tabla de dimensión proveedor:



Tablas de dimensionales de Sucursal y Producto las cuales son llenandas con los datos de compras y de ventas:



En el flujo de control hay dos flujos de datos más:



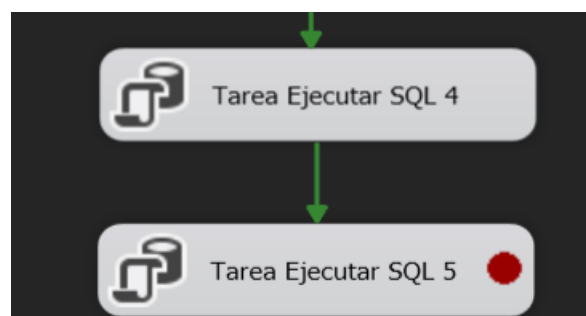
En el de llenado de compras se hace el proceso para unir las tablas dimensionales y relacionarlas a la tabla de hechos de compras usando la herramienta SSIS de “Combinación de mezcla” haciendo un inner join:



En el de llenado de ventas se hace el proceso para unir las tablas dimensionales y relacionarlas a la tabla de hechos de ventas usando la herramienta SSIS de “Combinación de mezcla” haciendo un inner join:



Por último en el flujo de control se encuentran otras dos herramientas SSIS “Tarea Ejecutar SQL”:





Las cuales tienen como única función guardar cuantos registros hay en las tablas del modelo datawarehouse y temporales:

Editor de la tarea Ejecutar SQL

Configure las propiedades requeridas para ejecutar instrucciones SQL y procedimientos almacenados mediante la conexión seleccionada.

General  
Asignación de parámetro  
Conjunto de resultados  
Expresiones

Nombre de variable	Dirección	Tipo de datos	Nombre de parámetro	Tamaño de parámetro
User::ContadorCompra	Output	NVARCHAR	0	-1
User::ContadorVenta	Output	NVARCHAR	1	-1
User::ContadorVendedor	Output	NVARCHAR	2	-1
User::ContadorProducto	Output	NVARCHAR	3	-1
User::ContadorSucursal	Output	NVARCHAR	4	-1
User::ContadorProveedor	Output	NVARCHAR	5	-1
User::ContadorCliente	Output	NVARCHAR	6	-1

Agregar Quitar

Aceptar Cancelar Ayuda

Editor de la tarea Ejecutar SQL

Configure las propiedades requeridas para ejecutar instrucciones SQL y procedimientos almacenados mediante la conexión seleccionada.

General  
Asignación de parámetro  
Conjunto de resultados  
Expresiones

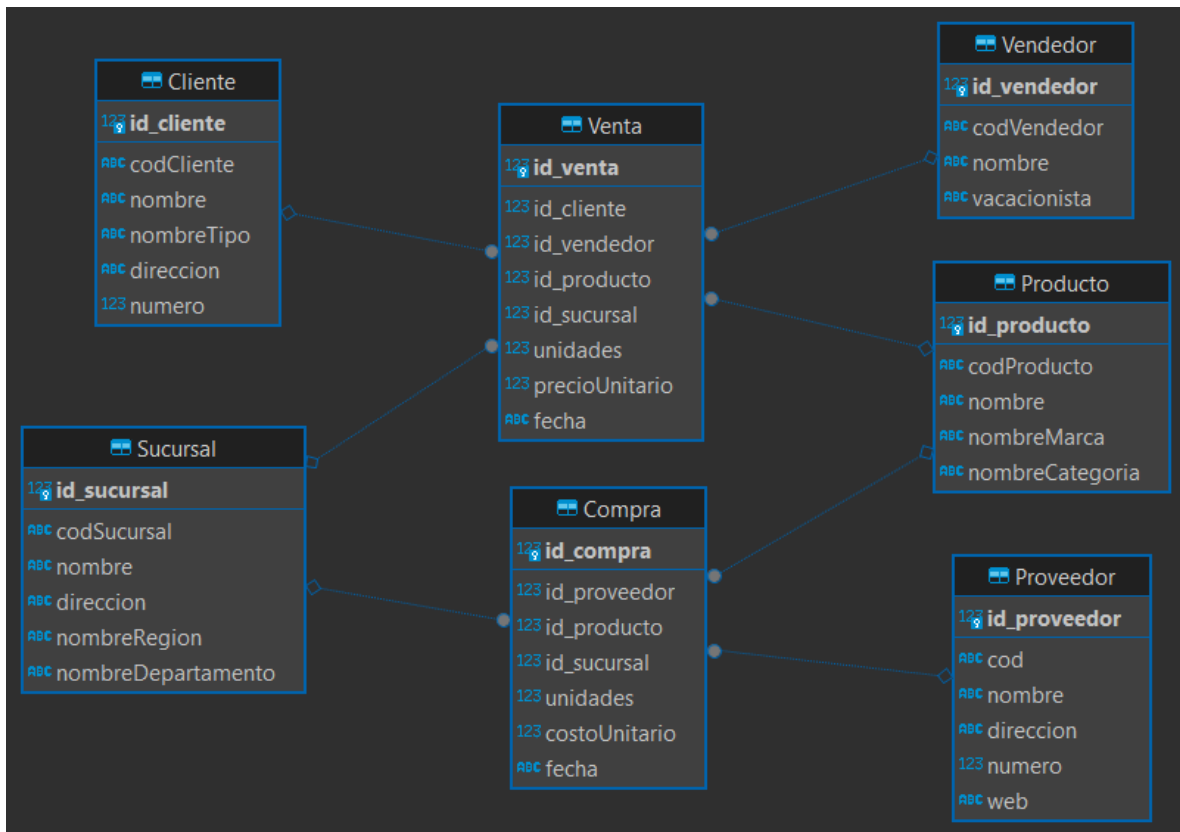
Nombre de variable	Dirección	Tipo de datos	Nombre de parámetro	Tamaño de parámetro
User::ContadorTempCompF2	Output	NVARCHAR	0	-1
User::ContadorTempVentF2	Output	NVARCHAR	1	-1

Agregar Quitar

Aceptar Cancelar Ayuda

## Modelo que se implementó para el DataWarehouse:

Por ultimo se tiene que el modelo de datawarehouse es el siguiente:



Se utilizó el modelo constelación donde hay dos tablas de hechos principales que son la de compra y venta ya que se facilita para hacer reportes de los mismos ya que comparten dimensionales como sucursal y producto el cual se adapta para ser utilizado de esta manera.