



# UNIVERSIDAD DON BOSCO

**Materia:** Datawarehouse y Minería de Datos

**Docente:** Ing. Karens Medrano

**Estudiante:** Eduardo Ezequiel López Rivera

**Carné:** LR230061

**Ciclo:** 01-2025

## Problemática

La Floristería **"Fiorella"** quiere saber cómo se compran sus productos y tiene la data de tres departamentos del país. Por ello, solicita su opinión sobre qué productos sobresalen, qué combinaciones son mejores y desea este estudio tanto por departamento como a nivel nacional.

Finalmente, los datos se cargarán en 4 archivos de Excel: **"Preferencias\_Departamentales.xlsx"**, **"Preferencias\_Nacionales.xlsx"**, **"Ventas\_Departamentales.xlsx"** y **"Ventas\_Nacional.xlsx"** incorporando el análisis de tendencias y patrones de compra.

## Contenido

Problemática .....	2
Configuración de proyecto en visual studio 2022 .....	4
<b><i>TOMAR EN CUENTA</i></b> .....	5
Solución .....	8
Desarrollo 10% .....	8
Desarrollo 25% .....	9
Desarrollo 40% .....	10
Desarrollo 50% .....	11
Desarrollo 60% .....	12
Desarrollo 75% .....	13
Desarrollo 85% .....	14
Desarrollo 100% .....	15
Ejecución exitosa.....	16

# Configuración de proyecto en visual studio 2022

En caso de ejecutar el proyecto, clonar el siguiente repositorio:

[https://github.com/eduardoezequieel/desafio1\\_DMD\\_EduardoLopez.git](https://github.com/eduardoezequieel/desafio1_DMD_EduardoLopez.git)

Existen 4 carpetas

- Fiorella\_Data\_CSV (Están contenidos los archivos .csv correspondientes a la data que se va a utilizar como origen).
- Fiorella\_Data\_Excel (Están contenidos los archivos .xlsx correspondientes a la data resultante de las operaciones, es decir, el destino).
- Integration Services Project1 (Contiene el proyecto de visual studio 2022).
- Moldes (Son las plantillas de los archivos en Excel, contiene únicamente las columnas, solo están de respaldo).

## **TOMAR EN CUENTA**

Crear base de datos, puede tener cualquier nombre, pero se deben crear las siguientes tablas:

```
CREATE TABLE [VentasFiorella] (  
    [id] nvarchar(50),  
    [RosasInt] int,  
    [ClavelesInt] int,  
    [MacetasInt] int,  
    [TierraInt] int,  
    [GirasolesInt] int,  
    [HortensialInt] int,  
    [GlobosInt] int,  
    [TarjetasInt] int,  
    [LiriosInt] int,  
    [AuroraInt] int,  
    [TulipanesInt] int,  
    [OrquideasInt] int,  
    [CarmesInt] int,  
    [ListonInt] int,  
    [Departamento] nvarchar(12)  
);
```

```
CREATE TABLE Combinaciones_Productos_Departamento (  
    Departamento NVARCHAR(255),  
    Producto1 NVARCHAR(255),  
    Producto2 NVARCHAR(255),  
    CantidadCompras INT  
);
```

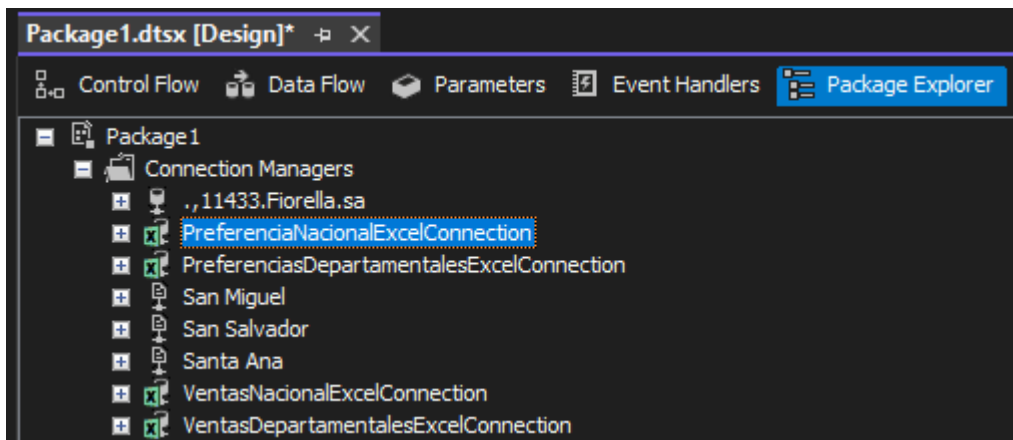
```
CREATE TABLE Combinaciones_Productos_Pais (

    Producto1 NVARCHAR(255),

    Producto2 NVARCHAR(255),

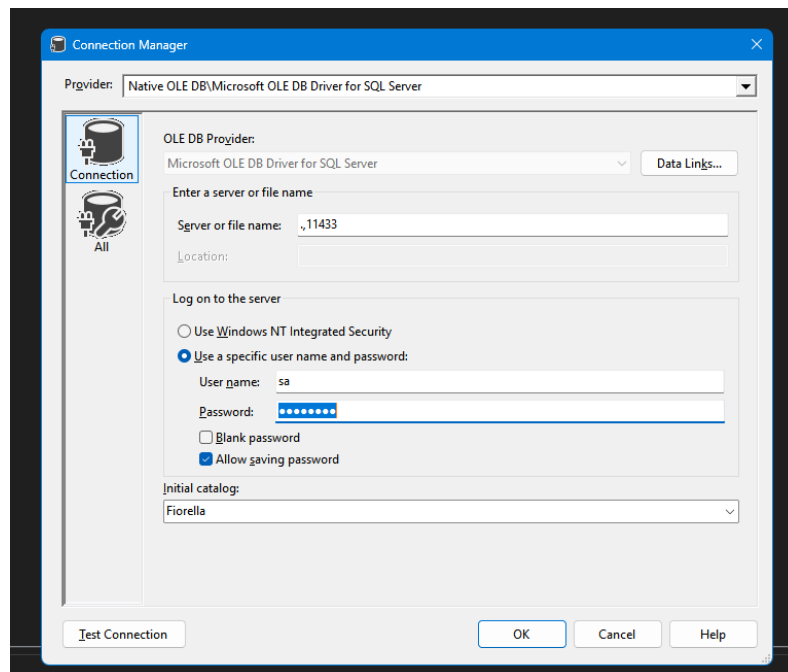
    CantidadCompras INT

);
```



Es necesario revisar cada una de las conexiones del proyecto y asignarlas a sus archivos correspondientes, ya que se usan rutas absolutas, estas no funcionarán justo después de clonar el proyecto. Se detalla a continuación cada una de las conexiones:

- **.,11433.Fiorella.sa:** Contiene la conexión hacia la base de datos en SQL Server, rellenar los campos a conveniencia y luego de verificar que funciona la conexión, por favor verificar que exista la base de datos con la tabla principal

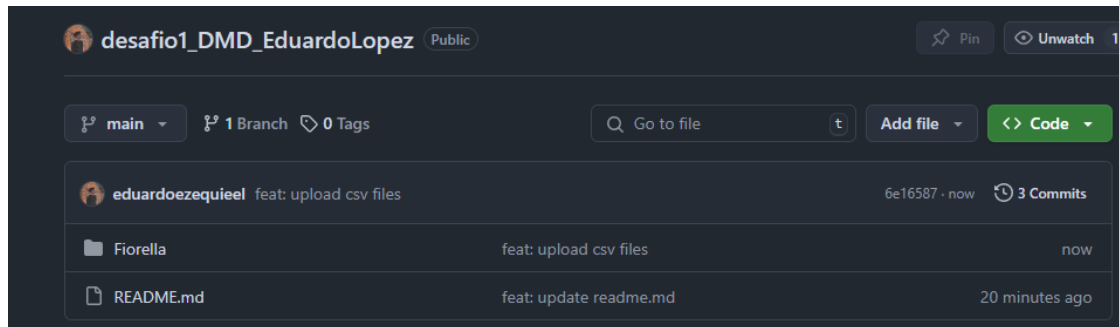


- **PreferenciaNacionalExcelConnection:** Ruta al archivo Preferencia\_Nacional.xlsx
- **PreferenciasDepartamentalesExcelConnection:** Ruta al archivo Preferencias\_Departamentales.xlsx
- **San Miguel:** Ruta al archivo .csv de las ventas en San Miguel.
- **Santa Ana:** Ruta al archivo .csv de las ventas en Santa Ana.
- **San Salvador:** Ruta al archivo .csv de las ventas en San Salvador.
- **VentasNacionalExcelConnection:** Ruta al archivo Ventas\_Nacional.xlsx
- **VentasDepartamentalesExcelConnection:** Ruta al archivo Ventas\_Departamentales.xlsx

# Solución

## Desarrollo 10%

Se crea un repositorio Git, alojado en GitHub para un control de versiones adecuado.





## Desarrollo 25%

Se agregan Flat File Sources, correspondiendo a cada una de las sucursales de la floristería, y se realizan conversiones de tipos de datos, las columnas de booleanos a integers, finalmente agregamos una nueva columna para identificar a que sucursal pertenece la venta.

Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.

Columns

Rosas

Claveles

Macetas

Tierra

Girasoles

Hortensia

Globos

Tarjetas

Lirios

Aurora

Tulpanes

Funciones matemáticas

Funciones de cadena

Funciones de fecha y hora

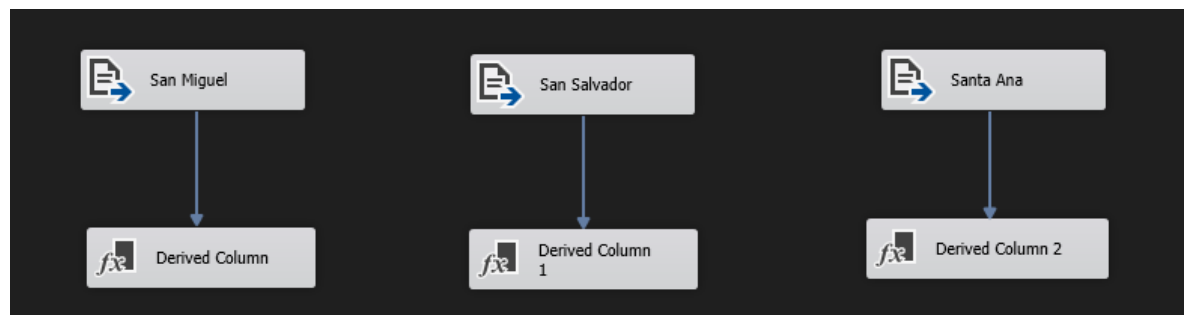
Funciones NULL

Conversiones de tipo

Operadores

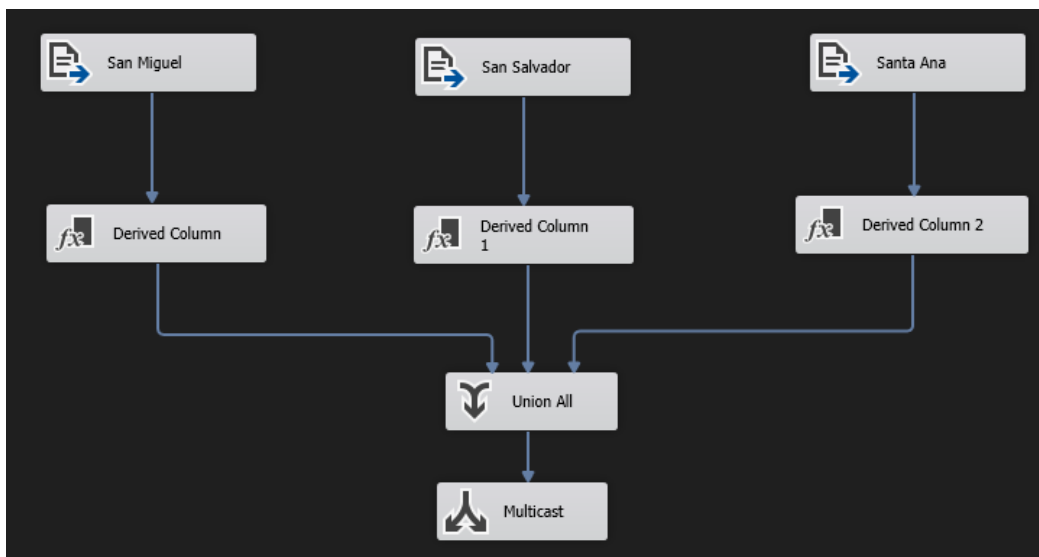
Descripción:

Derived Column Name	Derived Column	Expression	Data Type	Length	Precision	Scale	Code Page
RosasInt	<add as new column>	(Rosas ? 1 : 0)	four-byte signed integ...				
ClavelesInt	<add as new column>	(Claveles ? 1 : 0)	four-byte signed integ...				
MacetasInt	<add as new column>	(Macetas ? 1 : 0)	four-byte signed integ...				
Tierralnt	<add as new column>	(Tierra ? 1 : 0)	four-byte signed integ...				
GirasolesInt	<add as new column>	(Girasoles ? 1 : 0)	four-byte signed integ...				
Hortensialnt	<add as new column>	(Hortensia ? 1 : 0)	four-byte signed integ...				
GlobosInt	<add as new column>	(Globos ? 1 : 0)	four-byte signed integ...				
TarjetasInt	<add as new column>	(Tarjetas ? 1 : 0)	four-byte signed integ...				
LiriosInt	<add as new column>	(Lirios ? 1 : 0)	four-byte signed integ...				
Auroralnt	<add as new column>	(Aurora ? 1 : 0)	four-byte signed integ...				
TulpanesInt	<add as new column>	(Tulpanes ? 1 : 0)	four-byte signed integ...				
OrquideasInt	<add as new column>	(Orquideas ? 1 : 0)	four-byte signed integ...				
CarmesInt	<add as new column>	(Carmesi ? 1 : 0)	four-byte signed integ...				
ListonInt	<add as new column>	(Liston ? 1 : 0)	four-byte signed integ...				
Departamento	<add as new column>	"San Miguel"	Unicode string [DT_WS...	10			



## Desarrollo 40%

Unimos las 3 fuentes de datos con un **union all** y agregamos un **multicast** para poder tener varias salidas de datos



Output Column Name	Union All Input 4	Union All Input 1	Union All Input 2
Rosas	Rosas	Rosas	Rosas
Claveles	Claveles	Claveles	Claveles
Macetas	Macetas	Macetas	Macetas
Tierra	Tierra	Tierra	Tierra
Girasoles	Girasoles	Girasoles	Girasoles
Hortensia	Hortensia	Hortensia	Hortensia
Globos	Globos	Globos	Globos
Tarjetas	Tarjetas	Tarjetas	Tarjetas
Lirios	Lirios	Lirios	Lirios
Aurora	Aurora	Aurora	Aurora
Tulipanes	Tulipanes	Tulipanes	Tulipanes
id	id	id	id
Orquideas	Orquideas	Orquideas	Orquideas
Carmesi	Carmesi	Carmesi	Carmesi
Liston	Liston	Liston	Liston
RosasInt	RosasInt	RosasInt	RosasInt
ClavelesInt	ClavelesInt	ClavelesInt	ClavelesInt
MacetasInt	MacetasInt	MacetasInt	MacetasInt
Tierralnt	Tierralnt	Tierralnt	Tierralnt
GirasolesInt	GirasolesInt	GirasolesInt	GirasolesInt
Hortensialnt	Hortensialnt	Hortensialnt	Hortensialnt
GlobosInt	GlobosInt	GlobosInt	GlobosInt
TarjetasInt	TarjetasInt	TarjetasInt	TarjetasInt
LiriosInt	LiriosInt	LiriosInt	LiriosInt
Auroralnt	Auroralnt	Auroralnt	Auroralnt
TulipanesInt	TulipanesInt	TulipanesInt	TulipanesInt
OrquideasInt	OrquideasInt	OrquideasInt	OrquideasInt
CarmesiInt	CarmesiInt	CarmesiInt	CarmesiInt
ListonInt	ListonInt	ListonInt	ListonInt
Departamento	Departamento	Departamento	Departamento

## Desarrollo 50%

Se agregan dos **aggregate**, uno para calcular las ventas por producto por departamento y otro a nivel nacional. Luego aplicamos un **unpivot** para pasar las columnas de cada producto, a una sola columna llamada **Producto** y la suma de estas a una llamada **Cantidad**.

Input Column	Output Alias	Operation	Cc
RosasInt	Rosas	Sum	
ClavelesInt	Claveles	Sum	
MacetasInt	Macetas	Sum	
Tierralnt	Tierra	Sum	
GirasolesInt	Girasoles	Sum	
Hortensialnt	Hortensia	Sum	
GlobosInt	Globos	Sum	
TarjetasInt	Tarjetas	Sum	
LiriosInt	Lirios	Sum	
Auroralnt	Aurora	Sum	
TulipanesInt	Tulipanes	Sum	
OrquideasInt	Orquideas	Sum	
CarmesInt	Carmesi	Sum	
ListonInt	Liston	Sum	
Departamento	Departamento	Group by	

Input Column	Output Alias	Operation	Cc
RosasInt	Rosas	Sum	
ClavelesInt	Claveles	Sum	
MacetasInt	Macetas	Sum	
Tierralnt	Tierra	Sum	
GirasolesInt	Girasoles	Sum	
Hortensialnt	Hortensia	Sum	
GlobosInt	Globos	Sum	
TarjetasInt	Tarjetas	Sum	
LiriosInt	Lirios	Sum	
Auroralnt	Aurora	Sum	
TulipanesInt	Tulipanes	Sum	
OrquideasInt	Orquideas	Sum	
CarmesInt	Carmesi	Sum	
ListonInt	Liston	Sum	

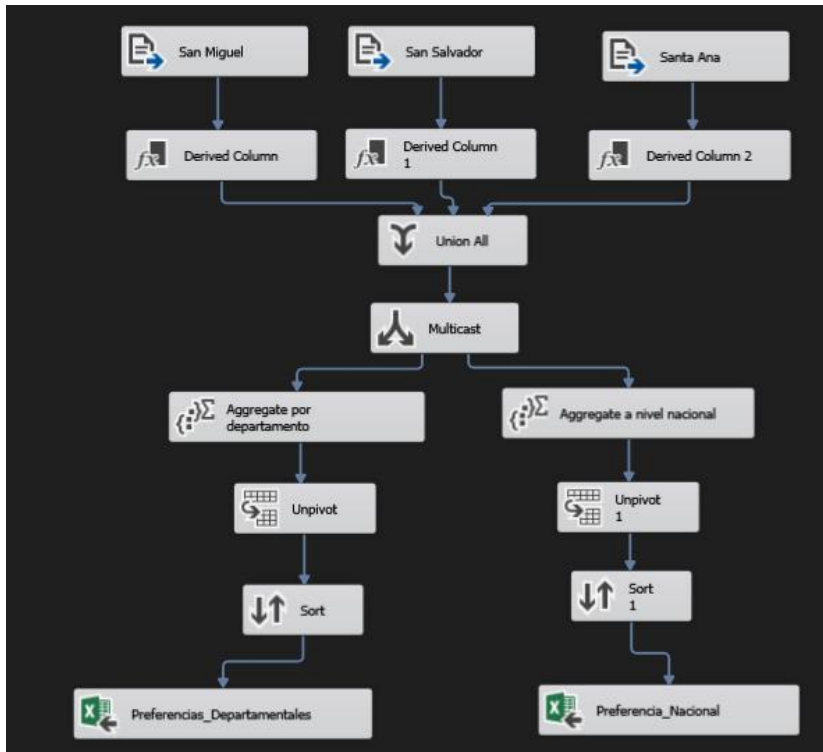
Input Column	Destination Column	Pivot Key Value
Hortensia	Cantidad	Hortensia
Globos	Cantidad	Globos
Tarjetas	Cantidad	Tarjetas
Lirios	Cantidad	Lirios
Aurora	Cantidad	Aurora
Tulipanes	Cantidad	Tulipanes
Orquideas	Cantidad	Orquideas
Carmesi	Cantidad	Carmesi

Pivot key value column name:

Producto

## Desarrollo 60%

Se agrega una salida para un archivo Excel que contendrá la información a nivel nacional y otro a nivel departamental, se crean las conexiones con los archivos que previamente poseen las columnas para que el modelado corresponda.



Excel Destination Editor

Configure the properties that enable the insertion of data via an Excel provider.

Connection Manager  
Mappings  
Error Output

Specify a connection manager, data source, or data source view for the Excel destination. Then, select the mode used to access data within the destination. After selecting the data access mode, select from among the additional data access options that appear.

Excel connection manager:  
PreferenciasDepartamentalesExcelConnection New...

Data access mode:  
Table or view

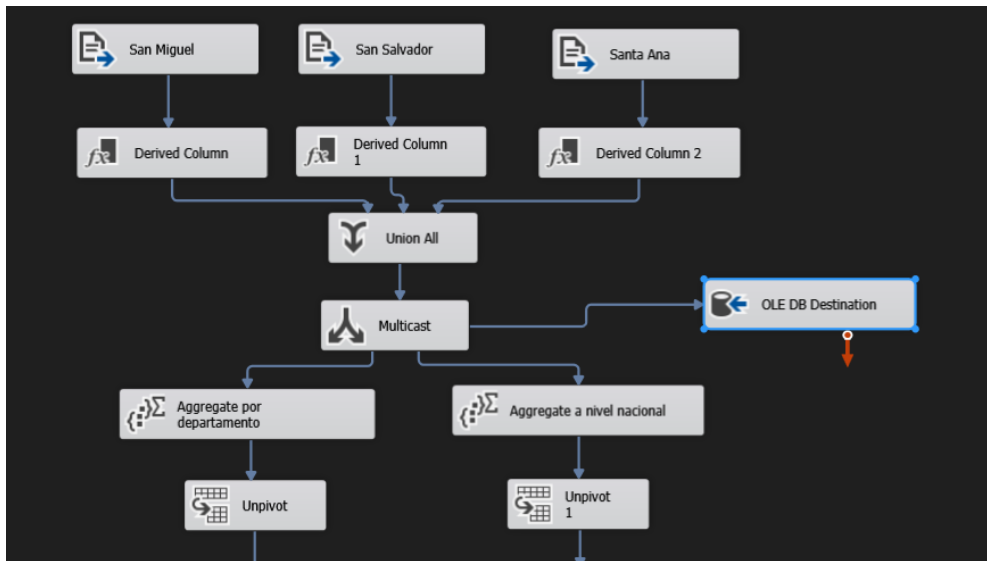
Name of the Excel sheet:  
Hoja1\$ New...

View Existing

OK Cancel Help

## Desarrollo 75%

Se agrega una tercera salida, en este caso un OLE DB Destination para guardar la data y poder hacer consultas mas avanzadas mas adelante. Es necesario asegurarnos de que exista la tabla de ventas (VentasFiorella)



OLE DB Destination Editor

Configure the properties used to insert data into a relational database using an OLE DB provider.

**Connection Manager**  
Mappings  
Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager:  
.,11433.Fiorella.sa New...

Data access mode:  
Table or view - fast load

Name of the table or the view:  
[VentasFiorella] New...

☐ Keep identity ☒ Table lock  
☐ Keep nulls ☐ Check constraints

Rows per batch:

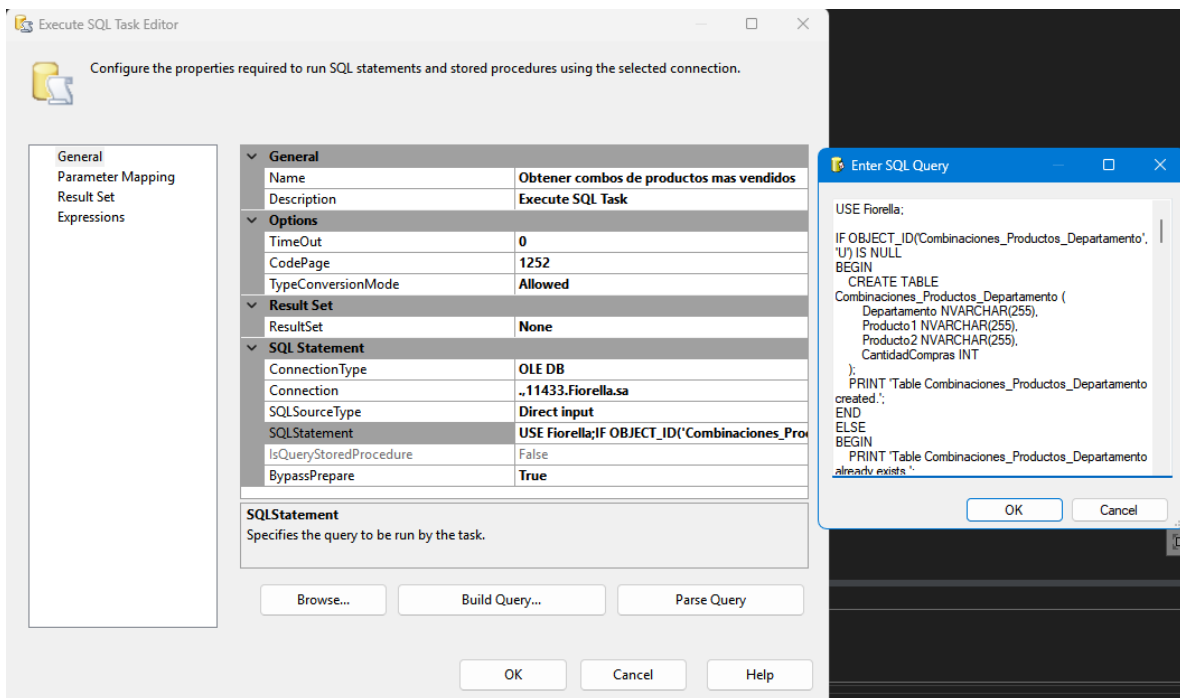
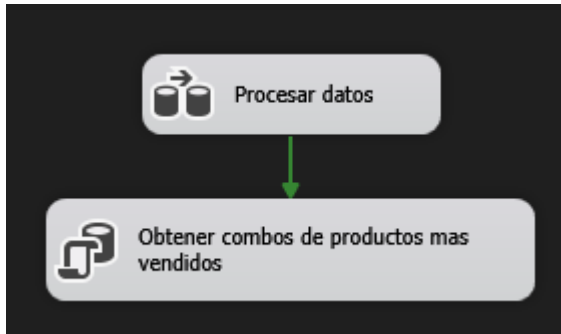
Maximum insert commit size:

View Existing

OK Cancel Help

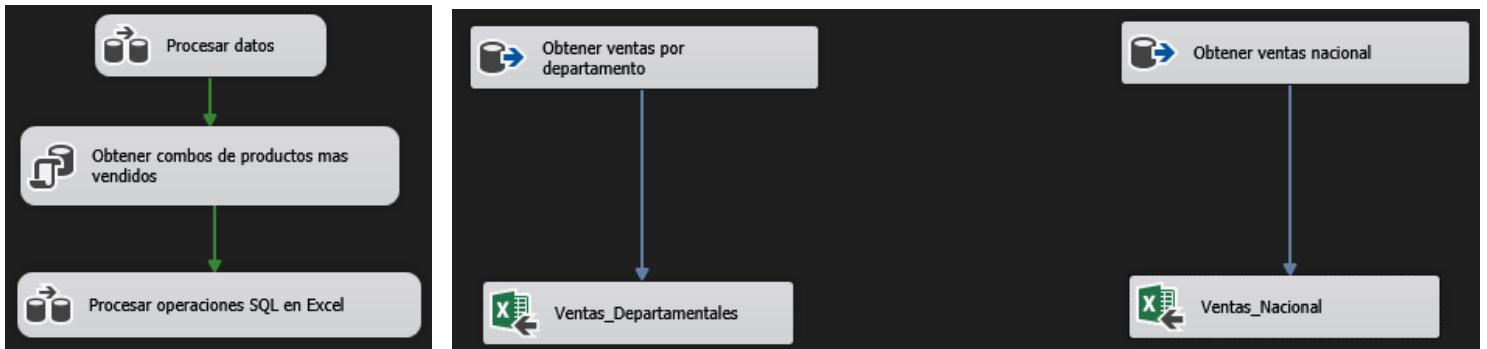
## Desarrollo 85%

Creamos un Execute SQL Task en nuestro Control Flow para hacer una consulta SQL y obtener los combos de productos mas vendidos a nivel departamental y nacional.



## Desarrollo 100%

Agregamos un data flow para leer los datos que se guardaron en el paso anterior y pasarlos los archivos correspondientes en formato .xlsx



OLE DB connection manager:

.,11433.Fiorella.sa

New...

Data access mode:

Table or view

Name of the table or the view:

[dbo].[Combinaciones\_Productos\_Departamento]

Excel connection manager:

VentasDepartamentalesExcelConnection

New...

Data access mode:

Table or view

Name of the Excel sheet:

Hoja1\$

New...

## Ejecución exitosa

