	UNIVERSIDAD DON BOSCO FACULTAD DE INGENIERIA ESCUELA DE COMPUTACION GUIA DE LABORATORIO N° 2
	Nombre de la práctica: Herramienta ETL , parte I Materia: Data WareHouse y Minería de Datos

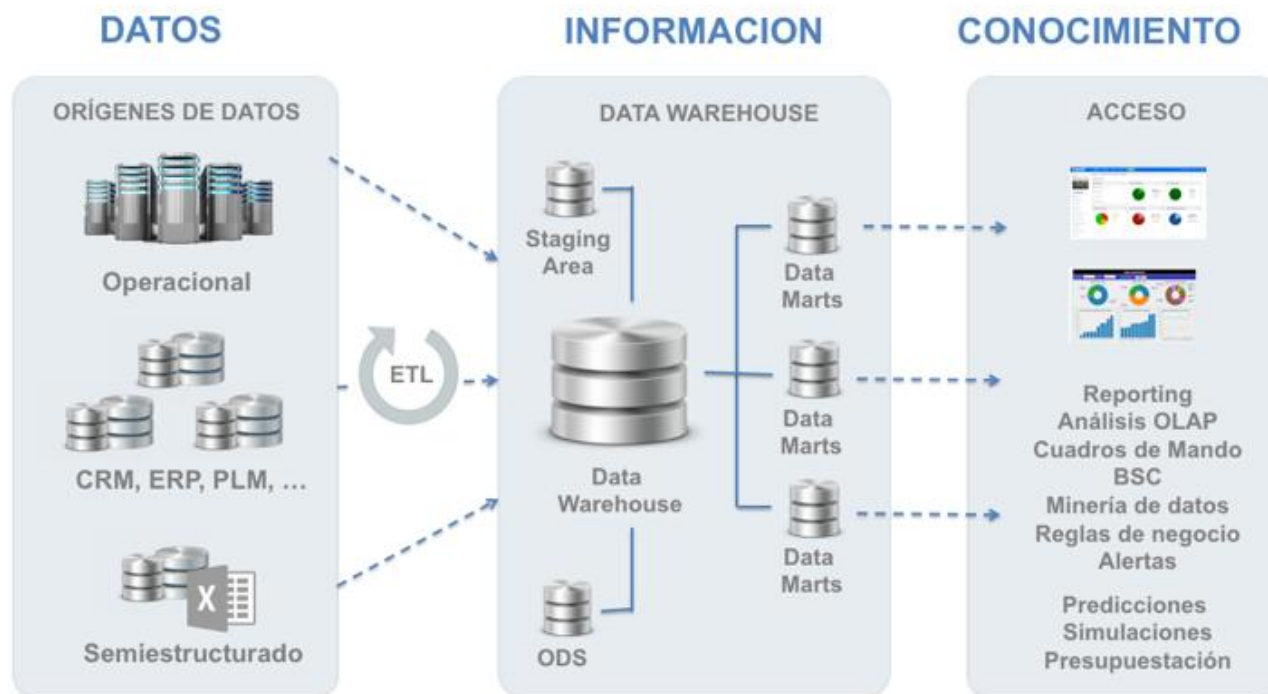
¿Qué es el Data Warehouse?

Un Data Warehouse es un gran almacén de datos e información que, además, recoge todos aquellos que son realmente necesarios para la realización de análisis e informe relacionado con el **Business Intelligence (BI)**. Una parte fundamental en la toma de decisiones de las grandes empresas a la hora de establecer objetivos, establecer normativas y plantear riesgos.

En todos los casos, el data warehouse cumple la función de guardar la información de interés para la empresa con fines prácticos y estratégicos. Sus estructuras determinan el grado de usabilidad, aunque sobre todo ayudan a saber hasta qué punto es más fácil orientarse a diferentes ramas de un mismo negocio en base a los datos obtenidos.

Para qué sirve el Data Warehouse

Aunque la finalidad principal del Data Warehouse es almacenar la información ya depurada para su posterior empleo a nivel empresarial, es una herramienta que facilita la toma de decisiones dentro de una compañía, así como también mejora la calidad de estas. Además, permite almacenar también consultas hechas anteriormente o análisis ya realizados, además de una mayor rapidez a la hora de acceder a la información. Por otra parte, es capaz de facilitar la comunicación entre departamentos en una misma compañía.



ETL (Extraction, Transformation, Load) se compone de tres fases principales:

- En la **extracción** se obtienen los datos de las fuentes de origen mediante descarga de ficheros planos de texto, o facilitados por el cliente, y luego se cargan en el repositorio (ODS) en tablas intermedias, que contienen los datos sin la estructura final del modelo.
- En la **transformación** se adecúa la información. En este proceso es típico duplicar tablas que contienen la información correcta y la creación de nuevos campos o nuevas tablas con datos agregados y/o calculados. Por ejemplo para agrupar información por criterios geográficos, temporales, o de estructura jerárquica o comercial que serán útiles para el análisis.
- **Carga de datos**, donde una vez reorganizada la información, la cargamos en las tablas definitivas de nuestro/s repositorio/s de datos: datawarehouse (corporativo) y/o datamart (departamental). Nuevamente se duplican las tablas que contienen la información correcta y posteriormente se crean los nuevos campos necesarios para contener toda la información.

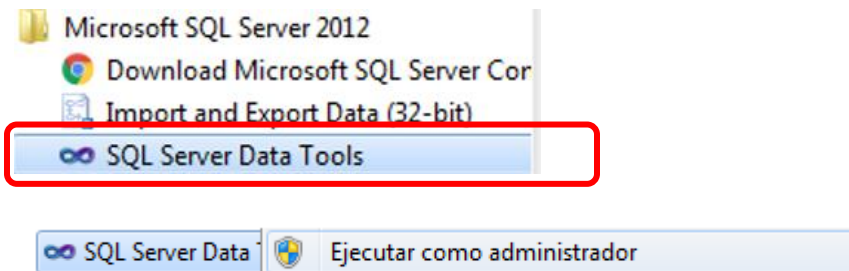
I. DESARROLLO

Ejercicio 1: Creación de un proyecto de Business Intelligence

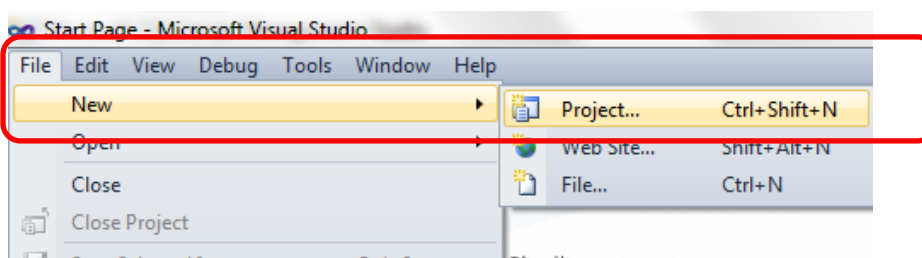
Antes de comenzar este ejercicio deberá crear:

- Una carpeta en el escritorio con el nombre: Proyecto_ETL_Carnet
- Dentro de la carpeta un archivo de Excel llamado Ejercicio1.xlsx

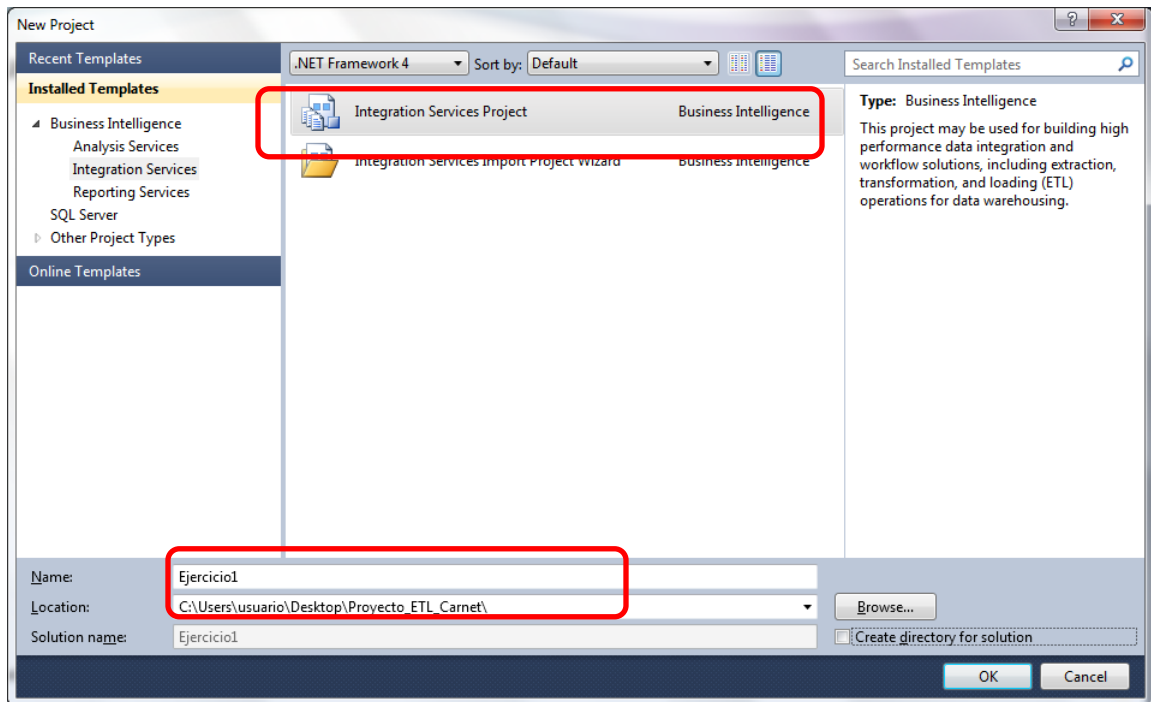
1. Abra **SQL Server Data Tools**, en modo administrador



2. Hacer clic en Business Intelligence
3. Hacer clic en Start Visual Studio
4. En el menú **Archivo (File)**, seleccione **Nuevo (New)** y haga clic en **Proyecto (Project)**.

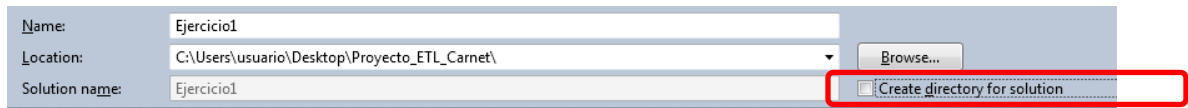


5. En el cuadro de diálogo Nuevo proyecto (New Project), en el panel Plantillas Instaladas (Installed Templates), seleccione el Proyecto de Integration Services o la plantilla del Asistente para proyectos de conexiones de Integration Services (Integration Services Import Project Wizard).

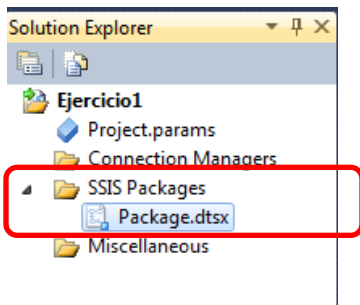


La plantilla que seleccione dependerá del tipo de proyecto que desee crear:

- La plantilla **Proyecto de Integration Services** crea un proyecto de Integration Services que contiene un único paquete vacío.
 - La plantilla Asistente para proyectos de conexiones de Integration Services también crea un proyecto de Integration Services que contiene un único paquete. Sin embargo, el paquete ya contiene los administradores de conexiones, orígenes de datos y destinos de datos que configuró en el asistente
6. En el cuadro **Nombre (Name)**, cambie el nombre predeterminado por **Ejercicio1**. Opcionalmente, desactive la casilla de verificación **Crear directorio para la solución**.



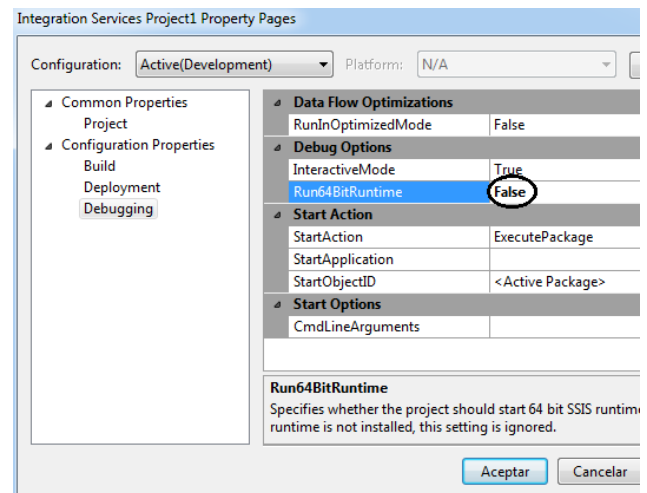
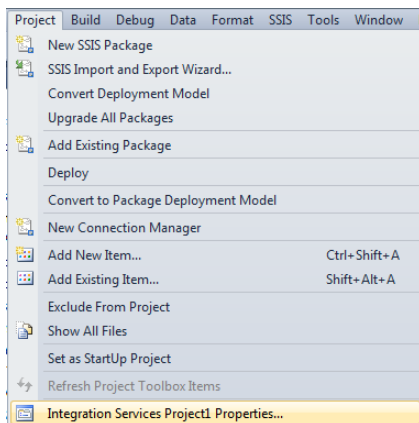
7. Acepte la ubicación predeterminada o haga clic en **Examinar (Browse)** para desplazarse a la carpeta que desee utilizar.
8. Haga clic en **Aceptar (OK)**.
9. De forma predeterminada, se creará un paquete vacío, denominado **Package.dtsx**, que se agregará al proyecto.



10. En la barra de herramientas del **Explorador de soluciones (Solution Explorer)**, haga clic derecho en **Package.dtsx**, haga clic en **Eliminar (Delete)** luego hacer en Aceptar (OK) para confirmar la eliminación

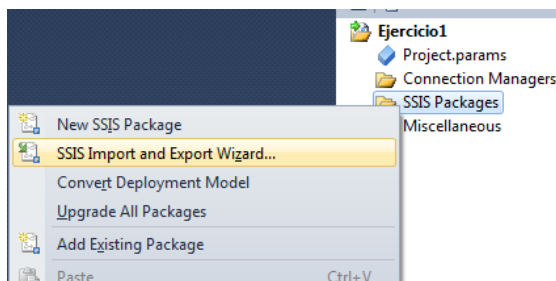
Quando se trabaja con sistemas de 64 bits, el proyecto puede dar errores al momento de la ejecución, para evitar este problema se realiza lo siguiente:

En el menú Project, seleccionar la última opción, (nombre de proyecto **properties**), en el cuadro de dialogo busque la opción **debugging** y modifique la propiedad **Run64BitRuntime** de True a **FALSE**



Crear un nuevo paquete SSIS

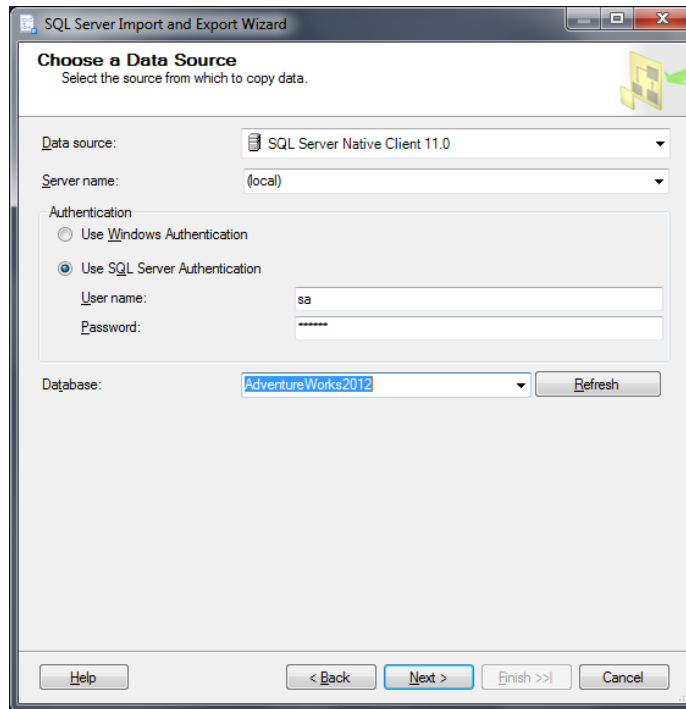
1. Dar clic derecho sobre la carpeta de paquetes SSIS (SSIS Packages) y hacer clic en Asistente para importación y exportación de SSIS (SSIS Import and Export Wizard)



2. En la opción origen de datos seleccione los siguientes valores:

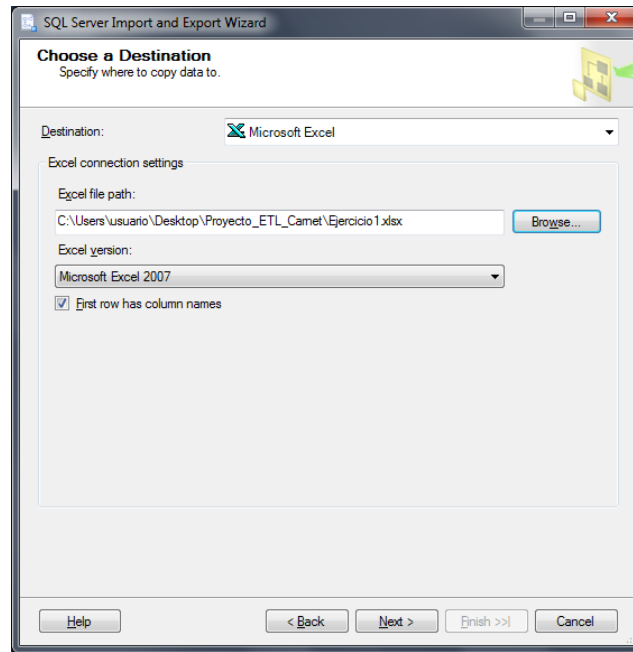
- Origen de datos: SQL Server Native Client 11.0
- Nombre de servidor: local o la dirección IP 127.0.0.1
- Autenticación SQL Server:
 - Usuario: sa
 - Contraseña: 123456
- Base de datos: AdventureWorks2012

Tal como se muestra en la siguiente figura:

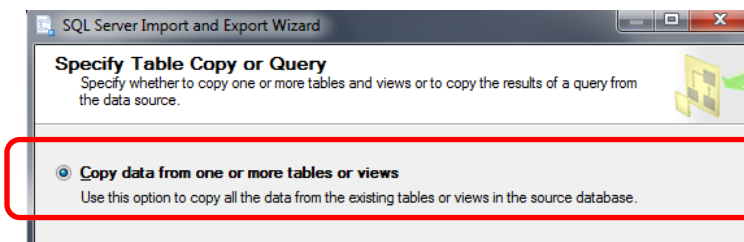


3. Dar clic en siguiente (Next)

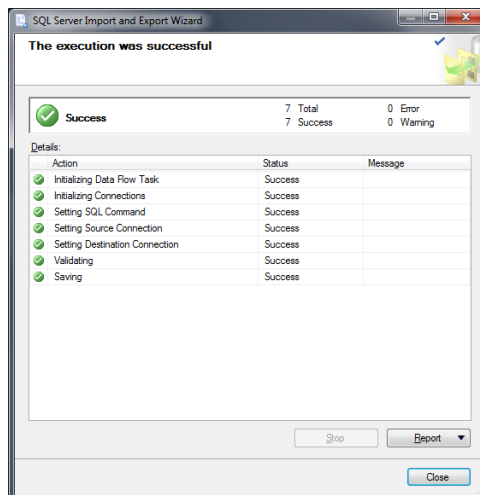
4. En la ventana Seleccionar Destino (Destination): **Microsof Excel**, en la ruta de acceso del archivo Excel, buscar el archivo con extensión XLS que creo anteriormente (Ejercicio1.xls), en la versión de Excel dejar el predeterminado, como se muestra a continuación.



5. En la siguiente ventana seleccionar: **Copiar los datos de una o varias tablas o vistas.**

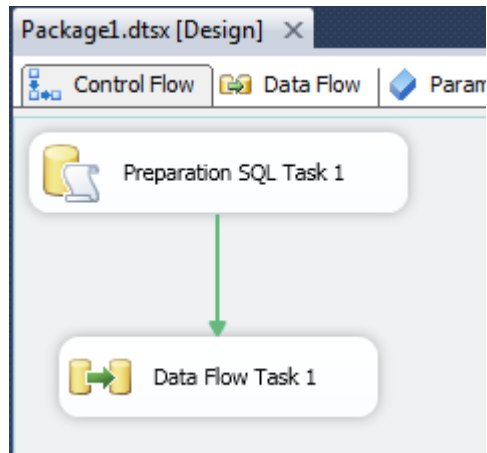


6. A continuación seleccionaremos la tabla (**HR Database**) HumanResources.Department
7. Clic en siguiente y luego clic en finalizar, y aparecerá la siguiente ventana en la que nos indicara si el proceso fue correcto



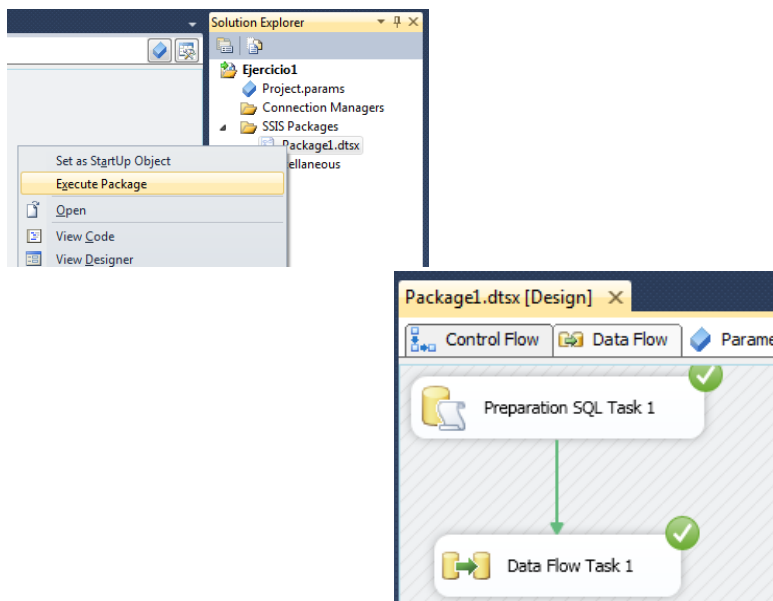
8. Dar clic en cerrar (Close)

Ha creado su primer paquete SSIS



Ejecutar el paquete SSIS

1. Ahora necesitamos Ejecutar el paquete, dar clic derecho sobre el paquete que acaba de crear, en la ventana del explorador de soluciones y seleccionar **Ejecutar Paquete (Execute Package)**

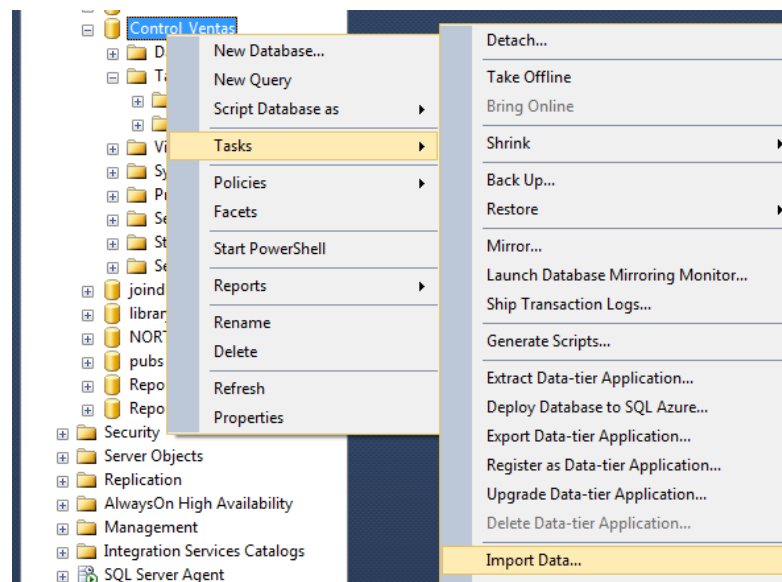


2. Cuando el proceso se halla ejecutado con éxito, abra el archivo .xlsx y observe las viñetas en la parte inferior del archivo.



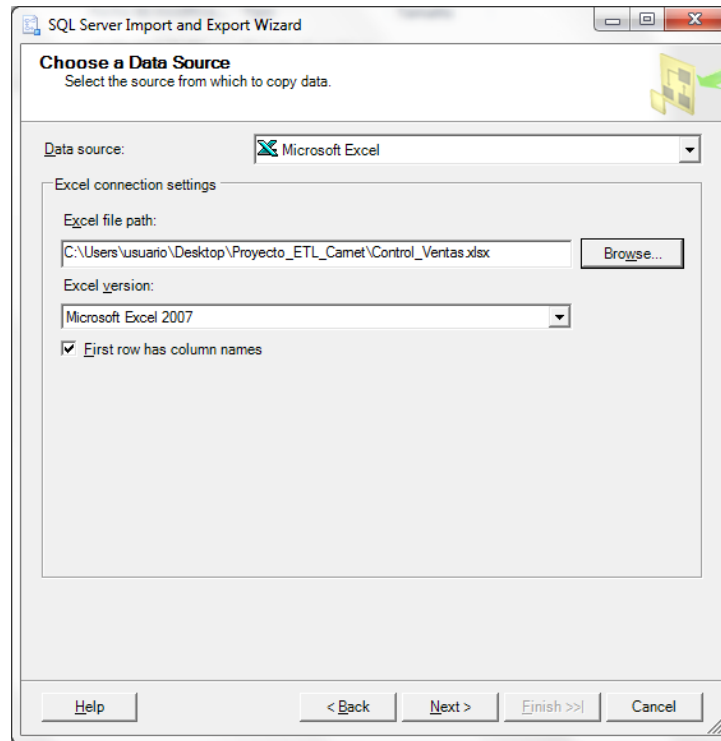
Ejercicio 2. Importar datos de un archivo .xlsx a SQL Server 2012

1. Debe bajar el archivo **Control_Ventas.xlsx**, del sitio de la Universidad (Recursos Guía #02 ETL) y guardarlo en la carpeta que se encuentra en el escritorio.
2. Entrar a **SQL Server Management Studio**
 - a. Hacer clic en el botón **Inicio**
 - b. Hacer clic en la opción **Todos los programas** y hacer clic en **Microsoft SQL Server 2012**
 - c. Hacer clic en SQL Server Management Studio
3. Para conectarse con el servidor de base de datos elija los siguientes parámetros de autenticación:
 - **Tipo de servidor:** Database Engine
 - **Nombre del servidor:** Colocar el nombre del servidor local, por ejemplo PCNumMaquina-SALA2
Nota: NumMaquina es el número de la maquina local
 - **Autenticación:** SQL Server Authentication
 - **Login:** sa
 - **Password:** 123456
4. Crear la base de datos **Control_Ventas**
5. Hacer clic derecho sobre el nombre de la base de datos, seleccionar la opción Tareas (Tasks) y hacer clic en Importar datos (Import Data...)



6. Se habilita el asistente para la importación y exportación de datos en SQL Server (**SQL Server Import and Export Wizard**)
7. En la ventana Seleccionar un origen de datos realizar los siguientes cambios:
 - Origen de datos: Seleccionar Microsoft Excel
 - Ruta de acceso del archivo de Excel: Buscar el archivo Control_Ventas.xlsx
 - En la opción Versión de Excel este seleccionada: Microsoft 2007 y que la casilla de verificación La primera fila tiene nombres de columna está activada.

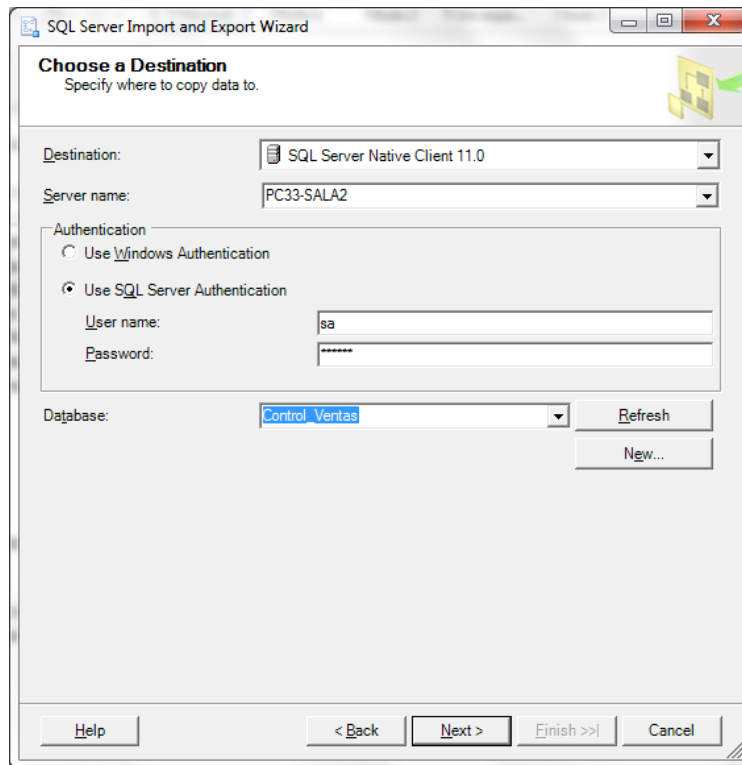
Así como se muestra en la siguiente figura:



8. Hacer clic en Siguiente (Next)

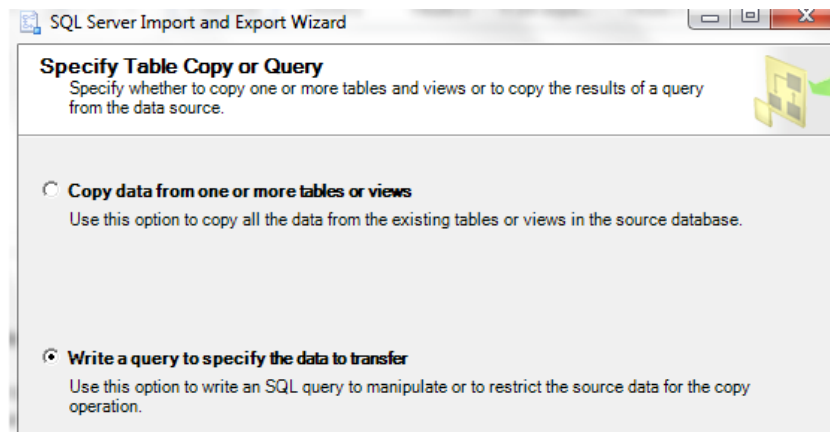
9. En la ventana Seleccionar un destino hacer las siguientes configuraciones:

- Destino: Seleccionar SQL Server Native Client 11.0
- Nombre del servidor: Seleccionar el servidor local
- Si la instancia del Motor de base de datos que ha especificado admite la autenticación de Windows, utilice el modo de **autenticación de Windows** predeterminado; en caso contrario, haga clic en Utilizar **autenticación de SQL Server** y escriba el nombre de usuario en el cuadro Nombre de usuario y la contraseña en el cuadro Contraseña (**Usuario: sa, Contraseña:123456**)
- En la lista Base de datos, seleccione **Control_Ventas** (La BD ya creada)
- Así como se muestra en la siguiente figura:



10. Hacer clic en Siguiente (Next).

11. En el paso Especificar copia de tabla o consulta (Specify Table Copy or Query) seleccionar la opción: Escribir una consulta para especificar los datos que se van a transferir (Write a query to specify the data to transfer)



12. Hacer clic en siguiente (next)

13. En la ventana Proporcionar una consulta de origen, escribir la siguiente instrucción SQL:

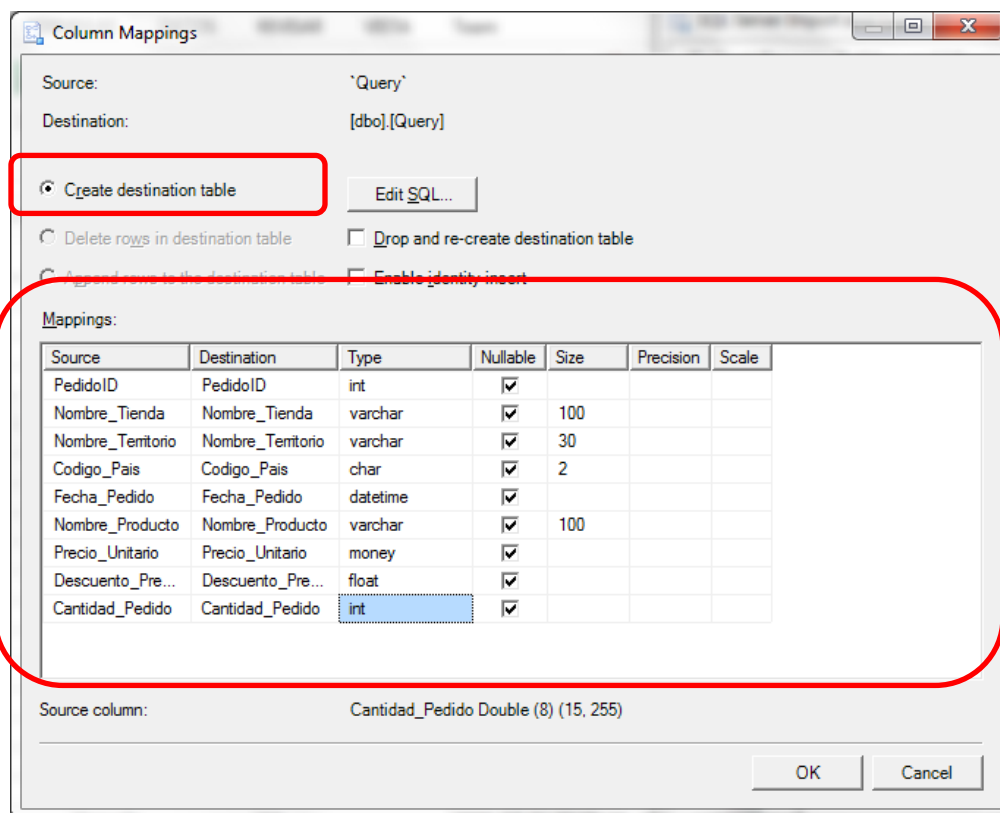
SELECT * FROM `Ventas`

Nota: los símbolos de `Ventas` no son comillas simples, son tildes invertidas ``

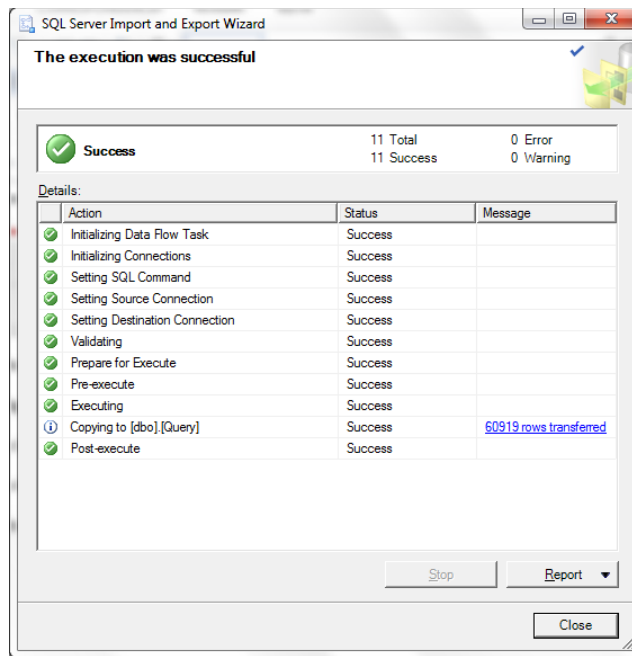
14. Hacer clic en Analizar (Parse) para verificar la sintaxis de la consulta



15. Hacer clic en Siguiente (Next)
16. Para modificar metadatos de columna y opciones de tabla, en la ventana Seleccionar tablas y vistas de origen (Select Source Tables and Views), haga clic en Editar asignaciones (Edit Mappings...).
17. En la ventana Asignaciones de columnas (Columns Mappings), compruebe que la opción Crear tabla de destino (Create destination table) está seleccionada, active la casilla de verificación Quitar y volver a crear la tabla de destino y modifique los metadatos de las columnas de destino.
18. Verifique que los cambios queden igual a como se muestra en la siguiente figura:

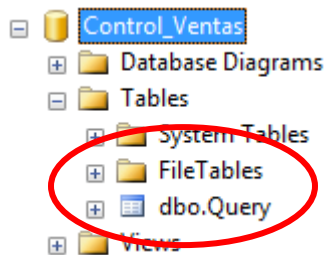


19. Haga clic en Aceptar (OK)
20. Haga clic en Siguiente (Next) (dos veces)
21. En la ventana Guardar y correr el paquete (Save and Run Package), hacer clic en siguiente
22. En la ventana Asistente completado (Complete the Wizard), hacer clic en Finalizar (Finish)
23. Debe de mostrar una ventana indicando la finalización correcta de la ejecución del paquete



24. Haga clic en Cerrar (Close)

25. Verifique en la Base de datos se ha creado la tabla dbo.Query



26. Realice un **SELECT** a la tabla y vera que contendrá los datos que tiene el archivo de Excel

FileEditViewQueryProjectDebugToolsWindowHelp

Control_Ventas

ExecuteDebug

Object Explorer

Connect

PC33-SALA2 (SQL Server 11.0.2100)

Databases

System Databases

Database Snapshots

AdventureWorks2012

AdventureWorksDW2012

Control_Ventas

Database Diagrams

Tables

System Tables

FileTables

dbo.Query

Views

Synonyms

Programmability

Service Broker

Storage

Security

joindb

library

NORTHWND

pubs

ReportServer

ReportServerTempDB

Security

Server Objects

SQLQuery2.sql - PC3...ol_Ventas (sa (54))

SELECT * FROM [dbo].[Query]

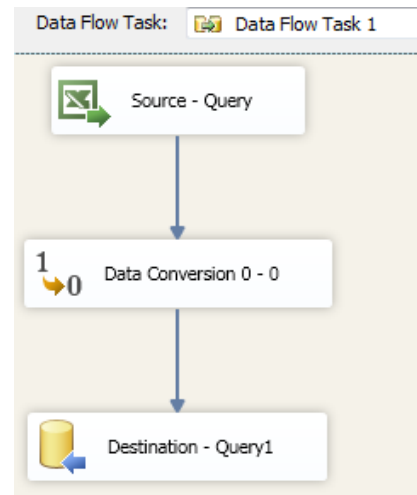
100 %

ResultsMessages

PedidoID	Nombre_Tienda	Nombre_Territorio	Codigo_Pais	Fecha_Pedido	Nombre_Producto	Precio_Unitario	
1	48306	Vigorous Exercise Company	Canada	CA	2006-12-01 00:00:00.000	Women's Tights, S	44.994
2	48306	Vigorous Exercise Company	Canada	CA	2006-12-01 00:00:00.000	Men's Sports Shorts, L	35.994
3	48306	Vigorous Exercise Company	Canada	CA	2006-12-01 00:00:00.000	AWC Logo Cap	5.1865
4	48306	Vigorous Exercise Company	Canada	CA	2006-12-01 00:00:00.000	Mountain-300 Black, 44	647.994
5	48306	Vigorous Exercise Company	Canada	CA	2006-12-01 00:00:00.000	ML Mountain Handlebars	33.7745
6	48306	Vigorous Exercise Company	Canada	CA	2006-12-01 00:00:00.000	ML Headset	61.374
7	48306	Vigorous Exercise Company	Canada	CA	2006-12-01 00:00:00.000	HL Mountain Frame - Black, 42	736.1455
8	48306	Vigorous Exercise Company	Canada	CA	2006-12-01 00:00:00.000	Long-Sleeve Logo Jersey, L	28.8404
9	48306	Vigorous Exercise Company	Canada	CA	2006-12-01 00:00:00.000	HL Fork	137.694
10	48306	Vigorous Exercise Company	Canada	CA	2006-12-01 00:00:00.000	Mountain-300 Black, 40	647.994
11	48306	Vigorous Exercise Company	Canada	CA	2006-12-01 00:00:00.000	LL Mountain Handlebars	24.2945
12	48306	Vigorous Exercise Company	Canada	CA	2006-12-01 00:00:00.000	Sport-100 Helmet, Black	20.1865
13	48306	Vigorous Exercise Company	Canada	CA	2006-12-01 00:00:00.000	ML Mountain Rear Wheel	141.615
14	48306	Vigorous Exercise Company	Canada	CA	2006-12-01 00:00:00.000	Sport-100 Helmet, Red	20.1865
15	48306	Vigorous Exercise Company	Canada	CA	2006-12-01 00:00:00.000	Long-Sleeve Logo Jersey, XL	28.8404
16	48306	Vigorous Exercise Company	Canada	CA	2006-12-01 00:00:00.000	Mountain-300 Black, 42	1736.4594

Ejercicio

1. Realice el mismo Ejercicio 2 pero creando un proyecto en la herramienta ETL (Integration Services Project) de SQL Server (SQL Server data tools)
2. Nombre del proyecto Ejercicio2
3. Utilizar el asistente para la importación de los datos, así como el Ejercicio 1
4. Utilice la misma base de datos como destino
5. Y el archivo Control_Ventas.xlsx como fuente de datos



Ejercicio 3. Crear un proyecto ETL sin el uso del asistente

Colocar en un archivo de Excel las ventas de los empleados que han realizado para Electrodomésticos y para Muebles. En SQL Server Management Studio realizar los siguientes pasos:

1. Crear la base de datos db_electrodomesticos en SQL Server

```
create database db_electrodomesticos;  
go  
use db_electrodomesticos;
```

2. Crear la tabla

```
create table ventas_electrodomesticos  
(  
    idempleado integer not null,  
    dui varchar(9) not null,  
    nit varchar(14) not null,  
    apellidos varchar(50) not null,  
    nombres varchar(50) not null,  
    ventas decimal(10,2) not null,  
    constraint pk_empleado primary key(idempleado)  
)
```

3. Insertar los siguientes registros

```
insert into ventas_electrodomesticos values(1,'0127859-5','0641140877-193','Hernandez  
Ramirez','Carlos Roberto',950.00)  
insert into ventas_electrodomesticos values(2,'1294587-6','7894120652-147','Perez  
Hernandez','Maria Leticia',758.25)  
insert into ventas_electrodomesticos values(3,'4589785-8','2365110578-145','Carranza','Pablo  
Jose',1050.20)  
insert into ventas_electrodomesticos values(4,'9658745-7','3658090976-145','Rodriguez  
Arias','Christian Alberto',857.50)  
insert into ventas_electrodomesticos values(5,'6985847-8','4758121280-193','Pastori','Xavier  
Iñaki',856.4)
```

```
select * from ventas_electrodomesticos
```

4. Crear la base de datos db_muebles en SQL Server

```
create database db_muebles  
go  
use db_muebles
```

5. Crear la tabla

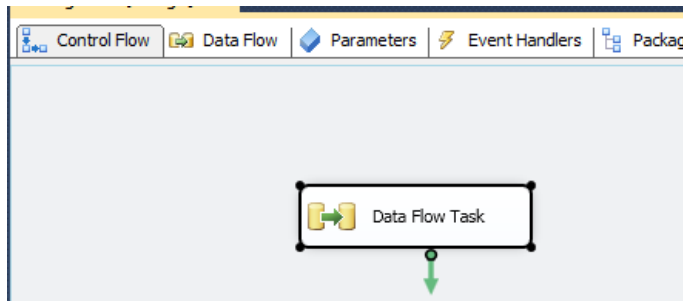
```
create table ventas_muebles  
(  
    idempleado integer not null,  
    dui varchar(9) not null,  
    nit varchar(14) not null,  
    apellidos varchar(50) not null,  
    nombres varchar(50) not null,  
    ventas decimal(10,2) not null,  
    constraint pk_empleado primary key(idempleado)  
)
```

6. Insertar los siguientes registros

```
insert into ventas_muebles values(1, '0127859-5', '0641140877-193', 'Hernandez Ramirez', 'Carlos Roberto', 750.00)
insert into ventas_muebles values(2, '1294587-6', '7894120652-147', 'Perez Hernandez', 'Maria Leticia', 958.25)
insert into ventas_muebles values(3, '4589785-8', '2365110578-145', 'Carranza', 'Pablo Jose', 850.20)
insert into ventas_muebles values(4, '9658745-7', '3658090976-145', 'Rodriguez Arias', 'Christian Alberto', 1057.50)
insert into ventas_muebles values(5, '6985847-8', '4758121280-193', 'Pastori', 'Xavier Iñaki', 1056.4)
```

```
select * from ventas_muebles
```

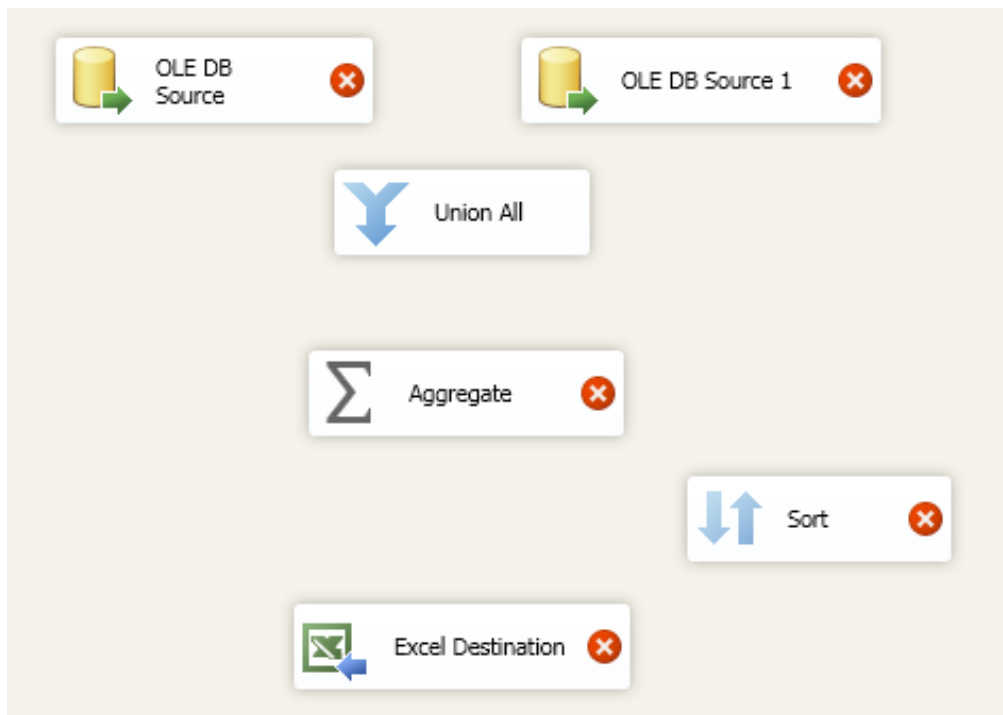
7. Crear un proyecto en la herramienta ETL (Integration Services Project) de SQL Server (SQL Server data tools)
8. Nombre del proyecto: **Ejercicio3**
9. Utilizar el paquete que viene por defecto en el `proyecto (Package.dtsx)
10. Agregar en la pestaña Control Flow un control **Data Flow Task**



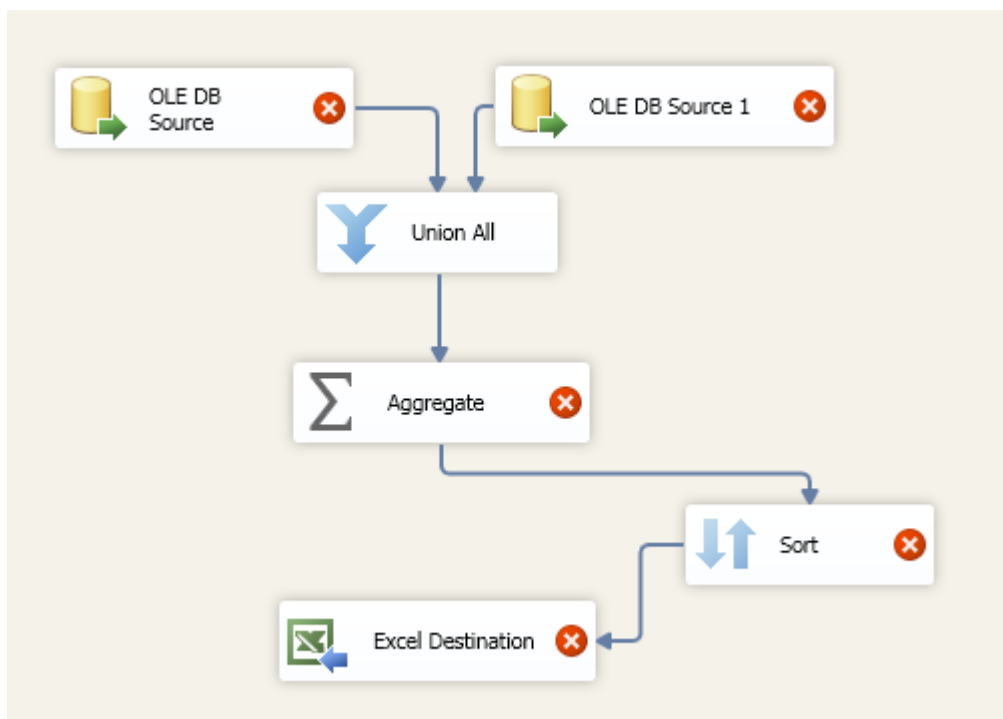
11. Hacer doble clic en el control y agregar los siguientes controles

Nombre del control	Cantidad
OLE DB Source	2
Union All	1
Aggregate	1
Sort	1
Excel Destination	1

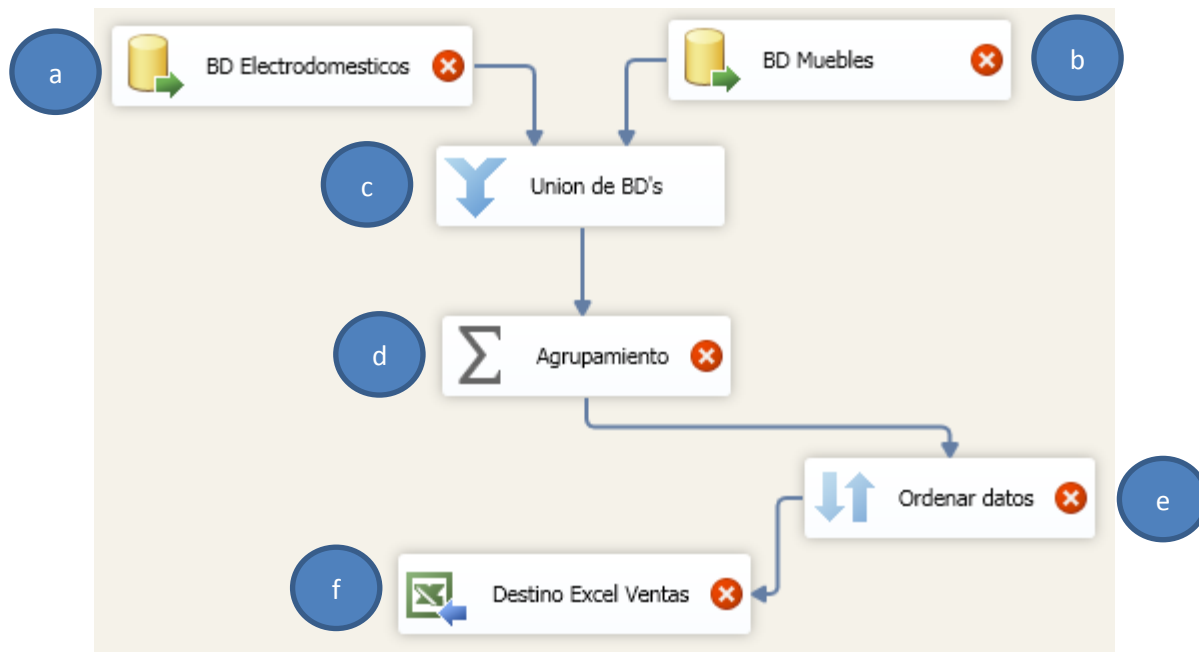
12. El paquete ETL queda de la siguiente manera:



13. Unir cada uno de los controles, así como se muestra a continuación:

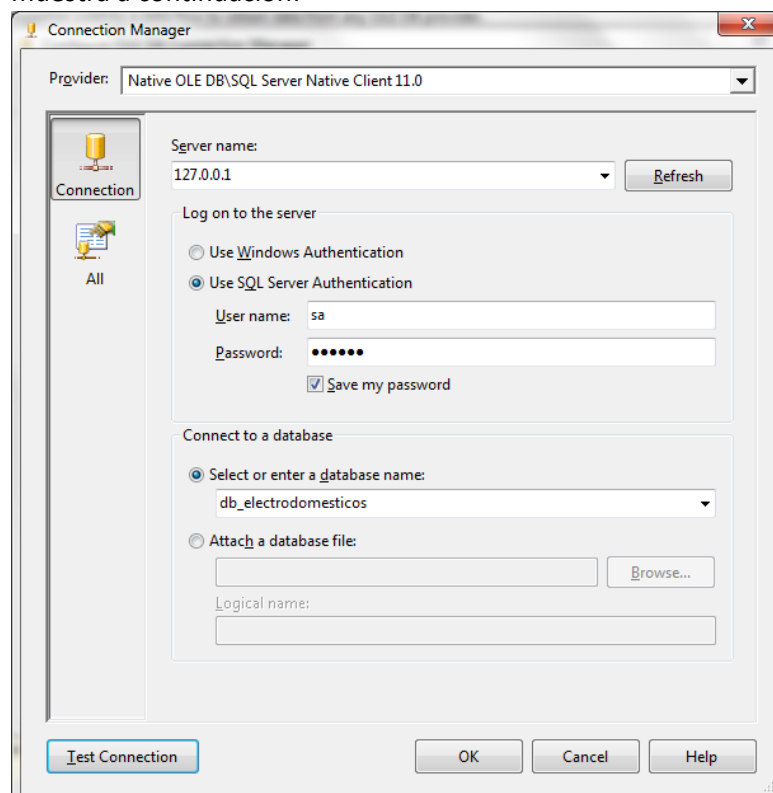


14. Realizar los siguientes cambios a cada control, hacer clic derecho y seleccionar la opción Rename, asignar a cada control los siguientes nombres así como se muestra a continuación:

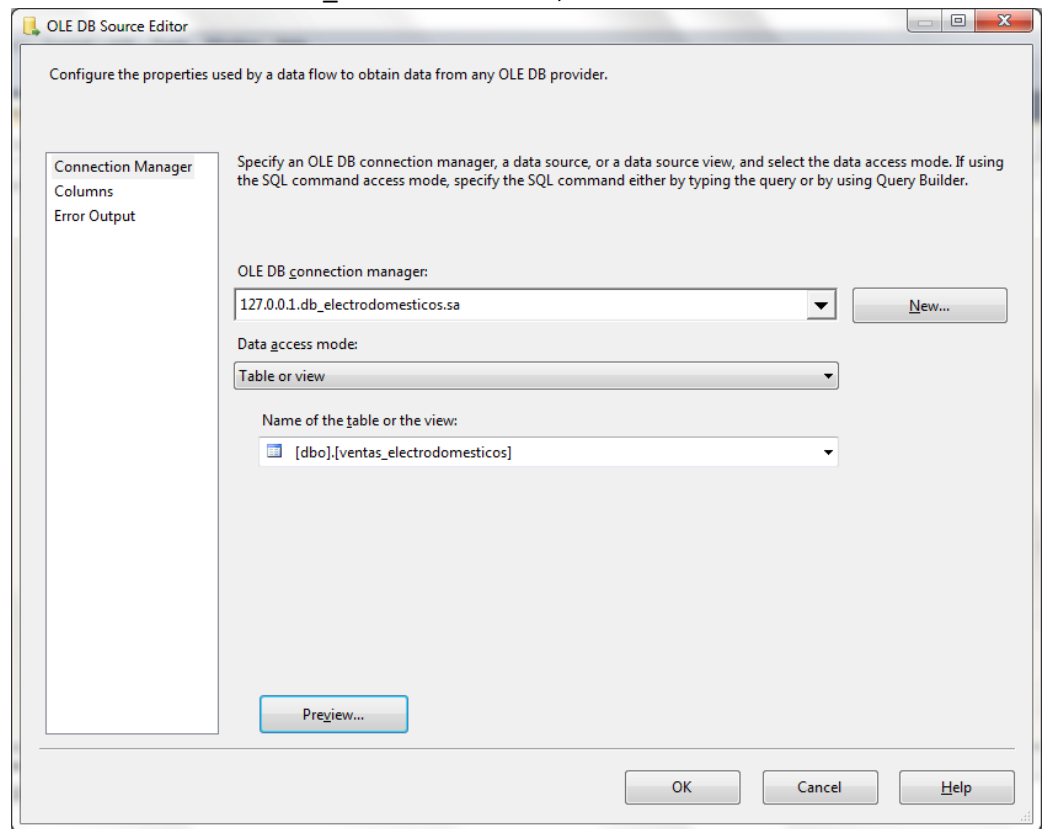


15. Agregar las siguientes configuraciones a cada control:

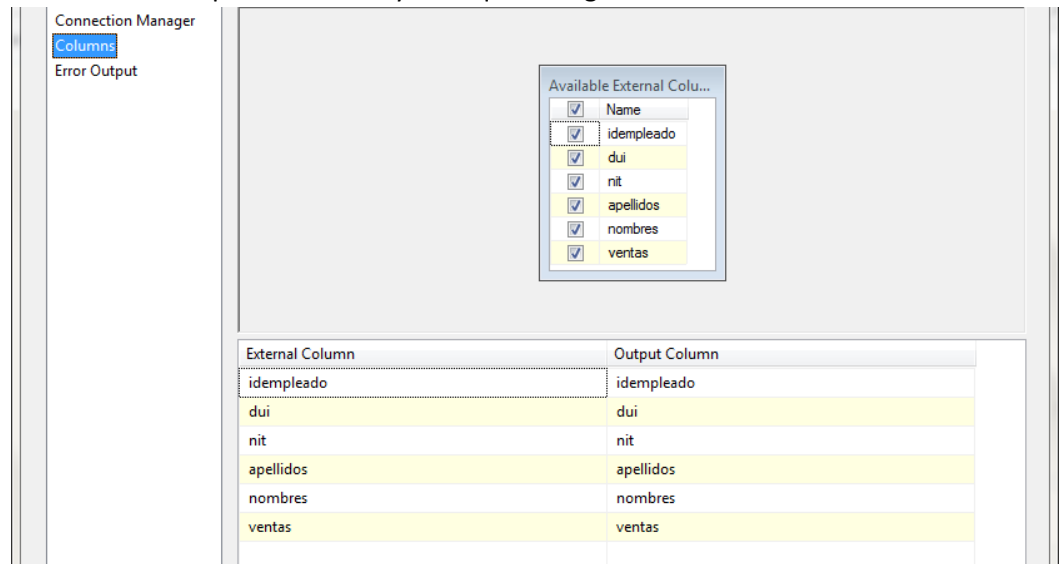
- a. Para la fuente de datos 1 (BD **Electrodomesticos**) seleccionar la base de datos db_electrodomesticos, siguiendo los pasos del asistente:
 - i. Hacer clic derecho sobre el control y seleccionar la opción Edit
 - ii. En la ventana OLE DB Source Editor, hacer clic en el botón New
 - iii. En la ventana Configure OLE DB Connection Manager, hacer clic en el botón New
 - iv. En la ventana Connection Manager, establecer las siguientes propiedad, así como se muestra a continuación:



- v. Probar la conexión, haciendo clic en Test Connection, si todo está bien hacer clic el botón OK.
- vi. Hacer clic en OK
- vii. Al estar en la ventana OLE DB Source Editor, en la opción Name of the table or the view seleccione la tabla: ventas_electrodomesticos, así como se muestra a continuación:

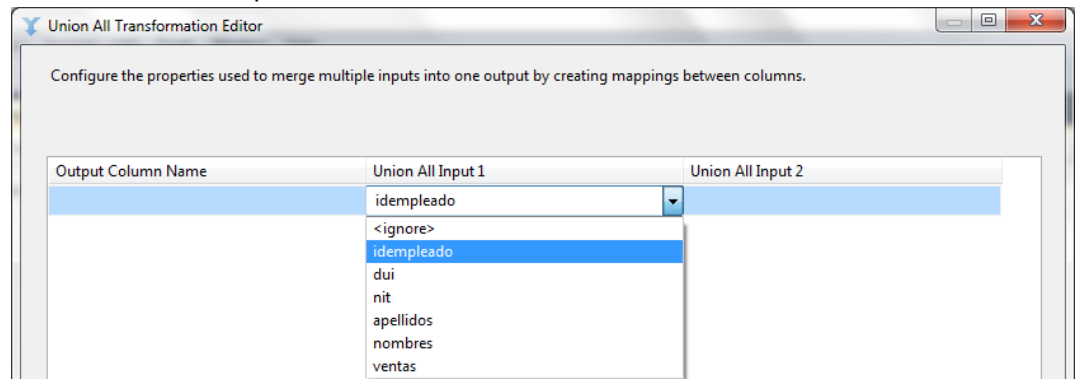


- viii. Hacer clic en la opción Columns y verifique las siguientes columnas de salida

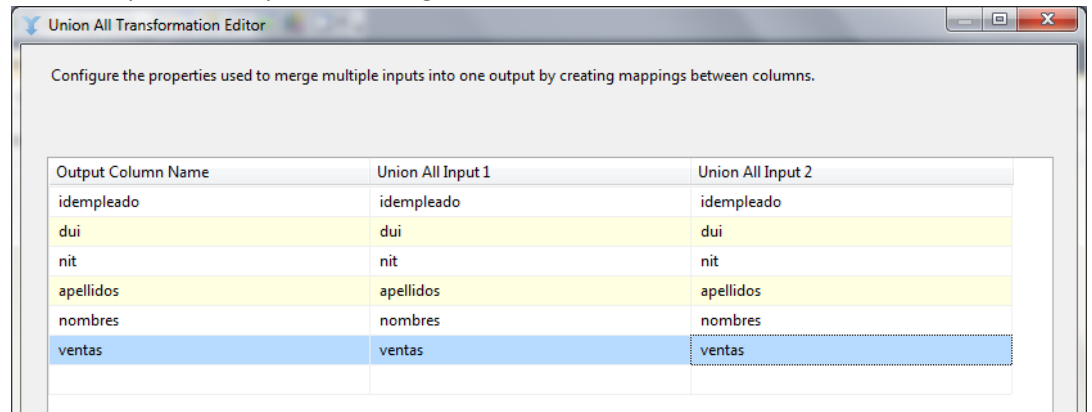


- ix. Hacer clic en el botón OK

- b. Para la fuente de datos 2 (BD Muebles) seleccionar la base de datos db_muebles y la tabla ventas_muebles, realizando los mismos pasos del punto a
- c. Para el control Union All (Union de BD's) hacer doble clic y realizar los siguientes pasos:
 - i. Para las columnas Union All Input 1 y Union All Input 2, debe ir seleccionando uno a uno cada uno de los campos de la tabla de la base de datos

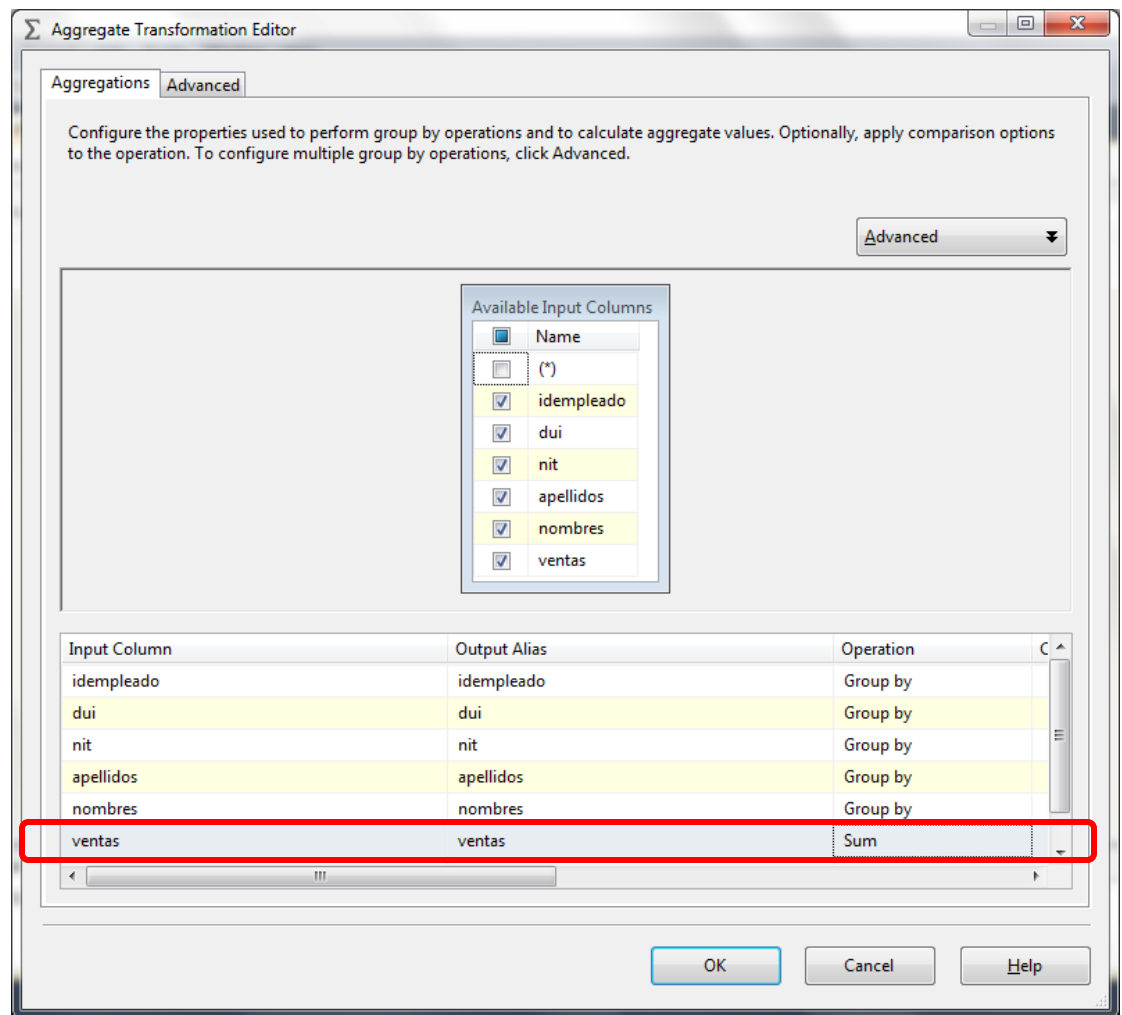


- ii. Verificar que al final queda de la siguiente manera:



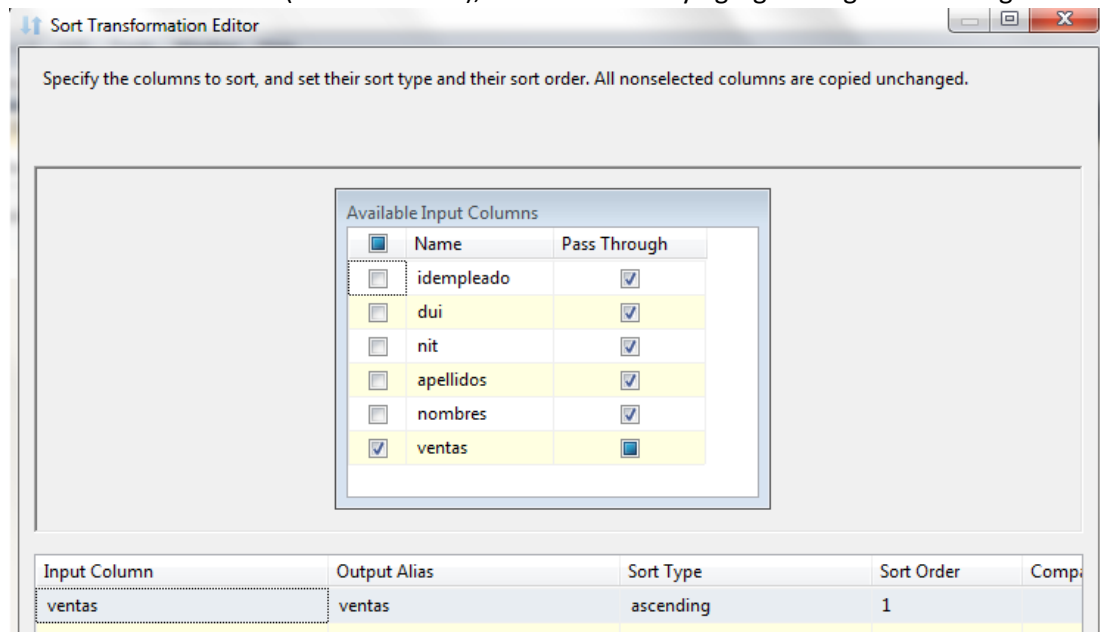
- iii. Hacer clic en OK

- d. Para el control Aggregate (Agrupamiento), hacer doble clic y verificar las siguientes propiedades:
 - i. Verificar que todos los campos estén agrupados (Group by)
 - ii. El campo ventas debe tener la función SUM

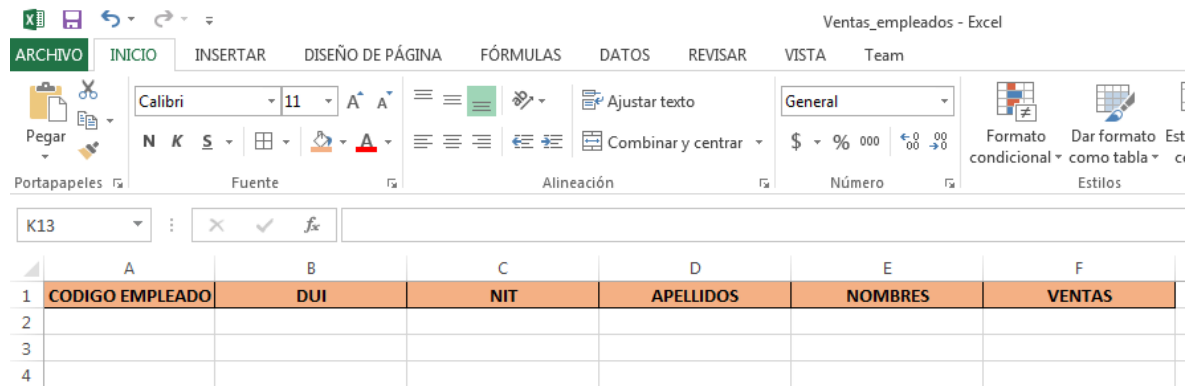


iii. Hacer clic en el botón OK

e. En la herramienta Sort (Ordenar datos), hacer doble clic y agregar la siguiente configuración



- f. En el destino Excel, debe seleccionar un archivo de Excel (nombre del archivo de Excel: Ventas_empleados.xlsx) el cual debe tener el siguiente formato:



Ventas_empleados - Excel

ARCHIVO INICIO INSERTAR DISEÑO DE PÁGINA FÓRMULAS DATOS REVISAR VISTA Team

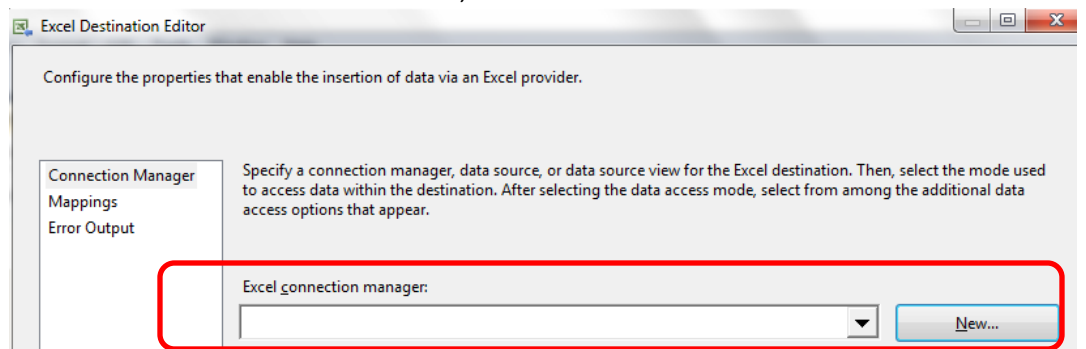
Calibri 11 Fuente Alineación Número Estilos

Portapapeles Pegar

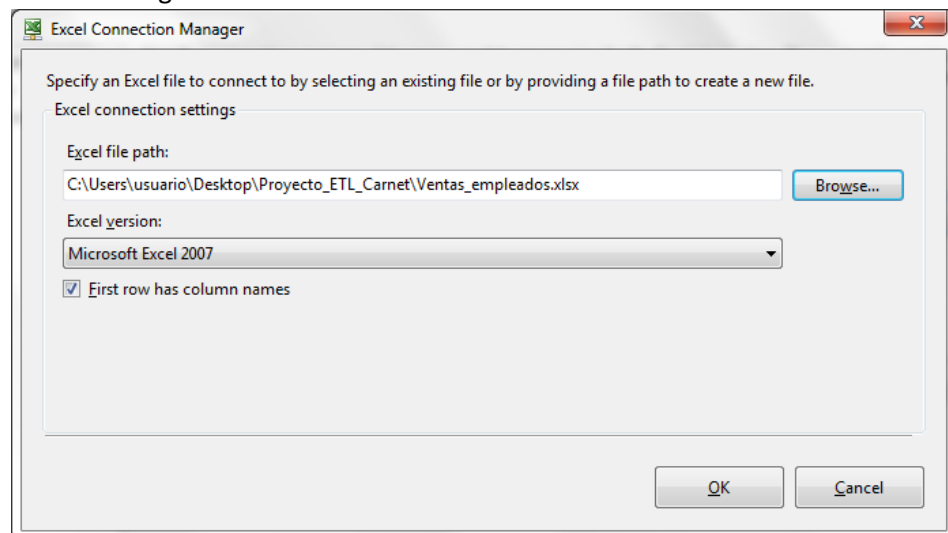
Formato condicional Dar formato Est

	A	B	C	D	E	F
1	CODIGO EMPLEADO	DUI	NIT	APELLIDOS	NOMBRES	VENTAS
2						
3						
4						

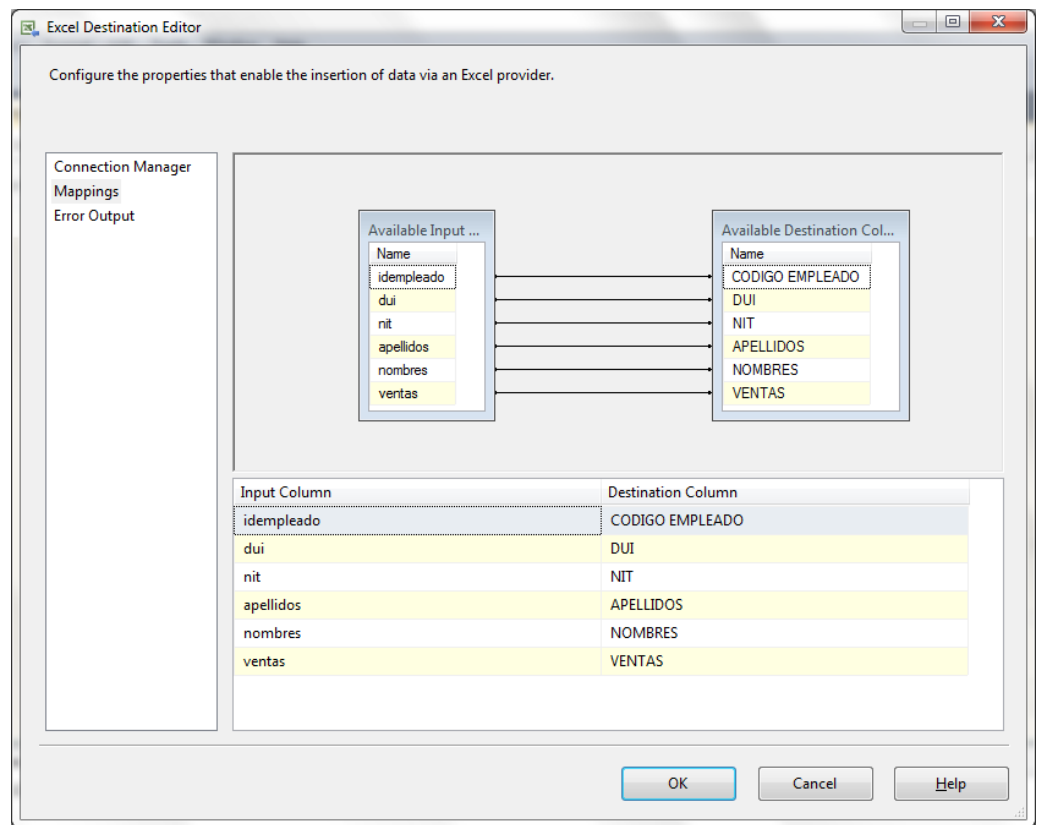
- i. Hacer doble clic sobre el control Excel Destination (Destino Excel Ventas)
- ii. En la ventana Excel Destination Editor, hacer clic en el botón New



- iii. Editar la siguiente ventana así como se muestra a continuación:

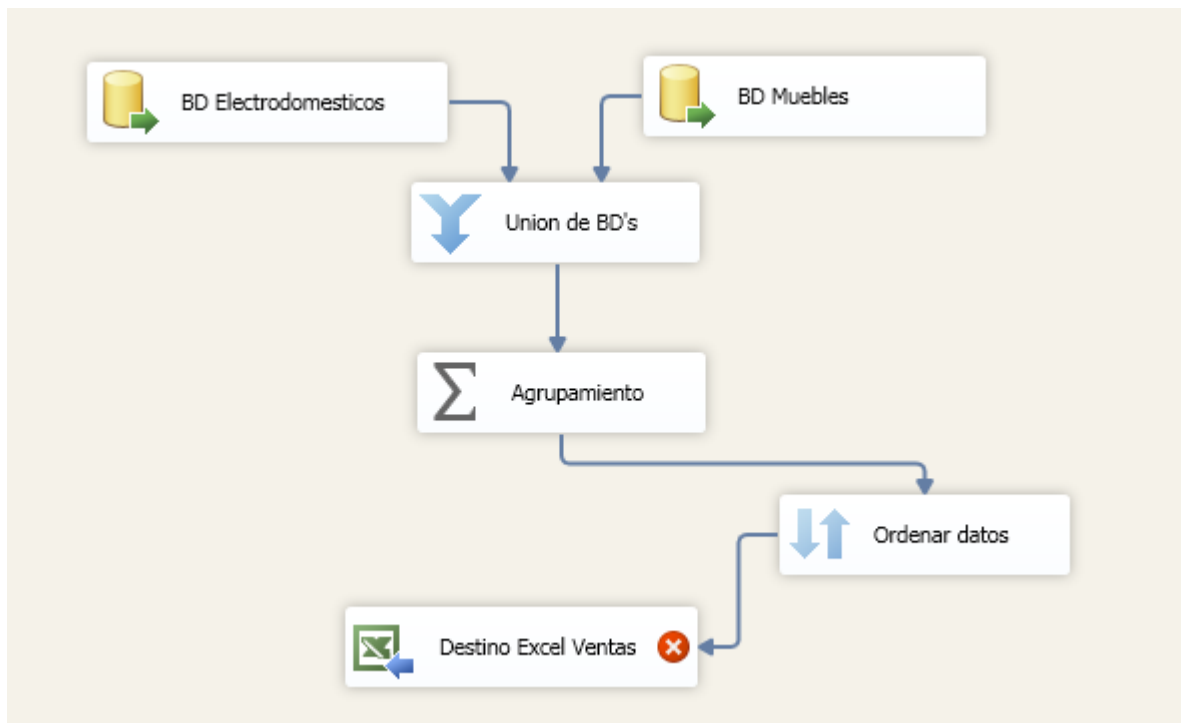


- iv. Hacer clic en el botón OK
- v. En la opción Name of the Excel sheet, seleccionar la Hoja1\$
- vi. Hacer clic en Mappings, verificar que la asignación se vea de la siguiente manera:



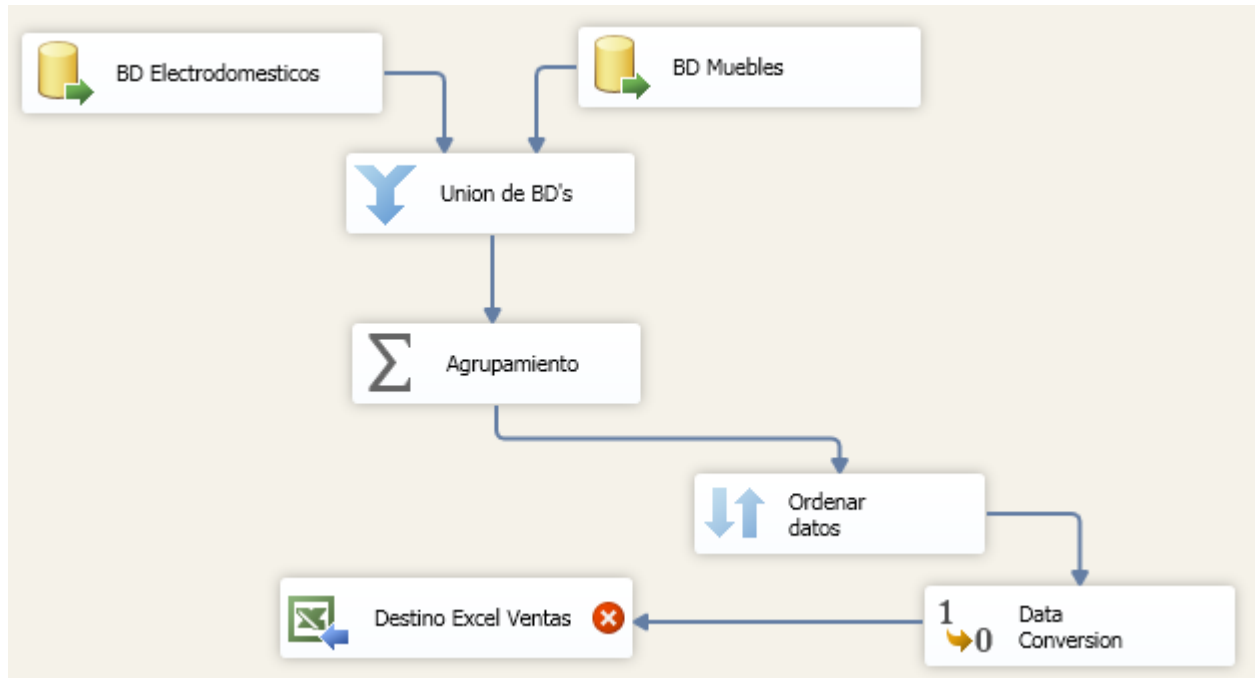
vii. Hacer clic en el botón OK

16. Al final su paquete ETL queda de la siguiente manera:

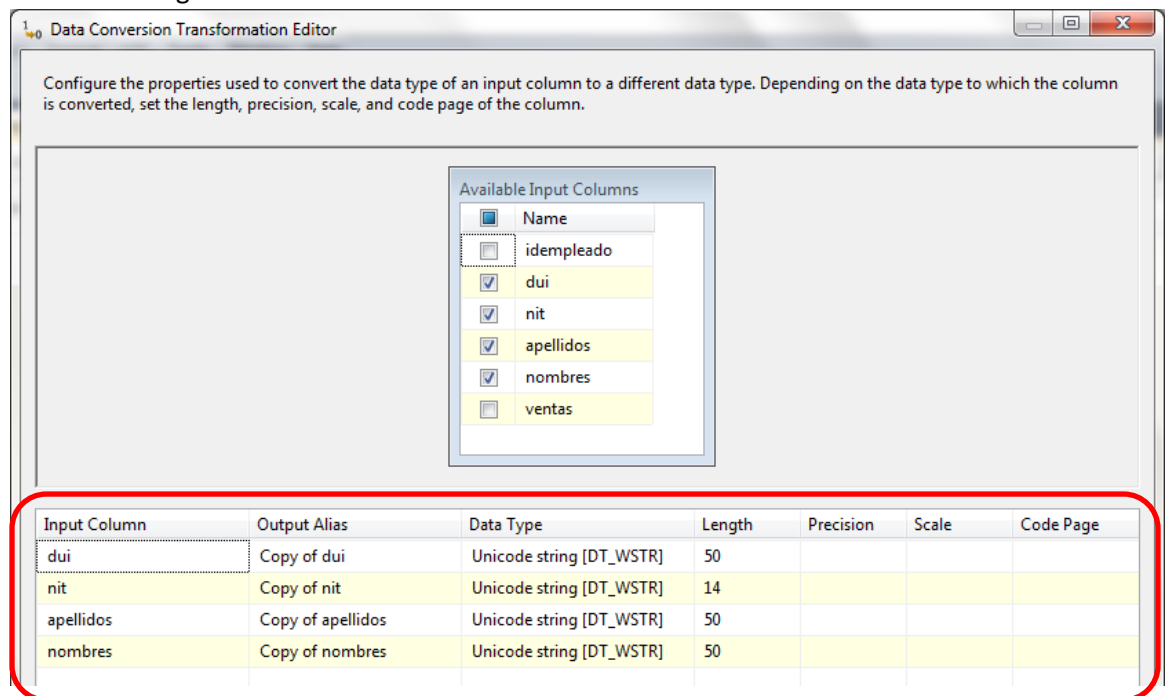


Si se da cuenta todavía muestra un error el control de destino (Destino Excel Ventas), al verificar el error (pasando el puntero sobre el control) coloca el siguiente mensaje: **Columns “dui” and “DUI” cannot convert between Unicode and non-unicode string data types**

Entonces se debe agregar un control de conversión de datos: **Data Conversion**, el paquete debe quedar de la siguiente manera:

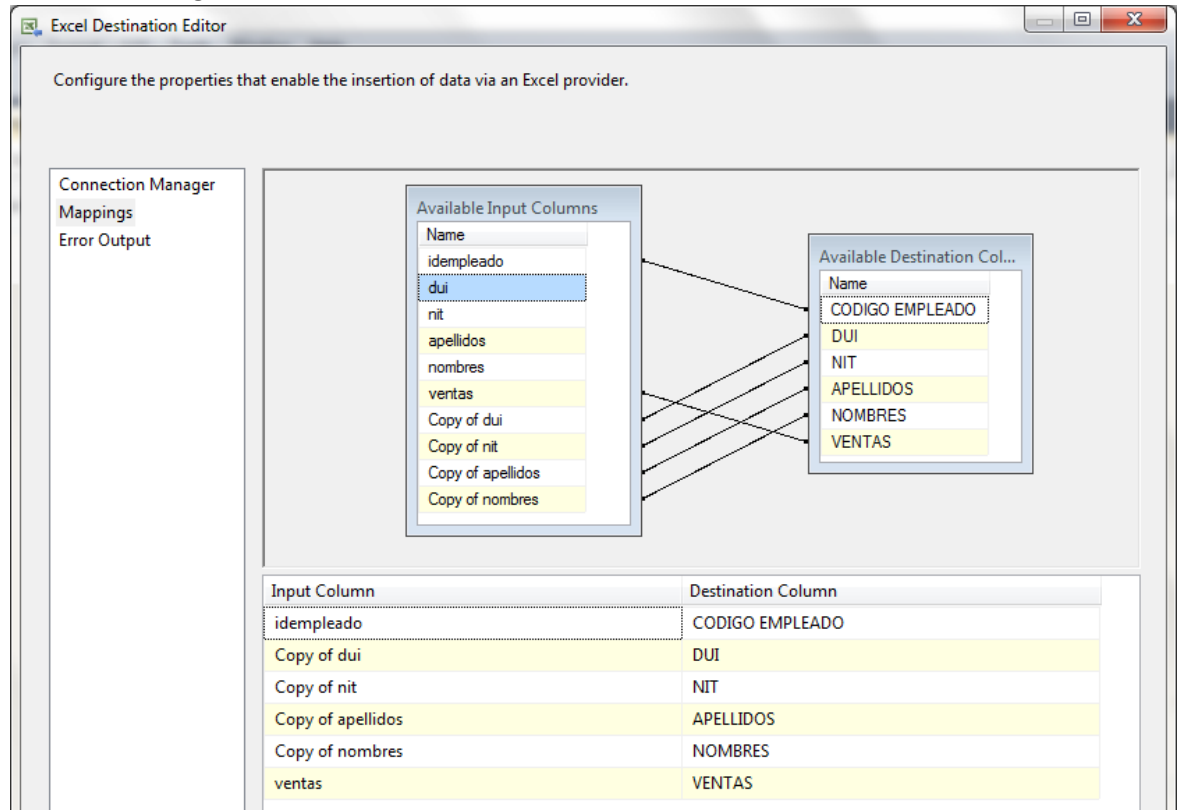


- i. Hacer doble clic en el control Data Conversion
- ii. Verificar los siguientes cambios:



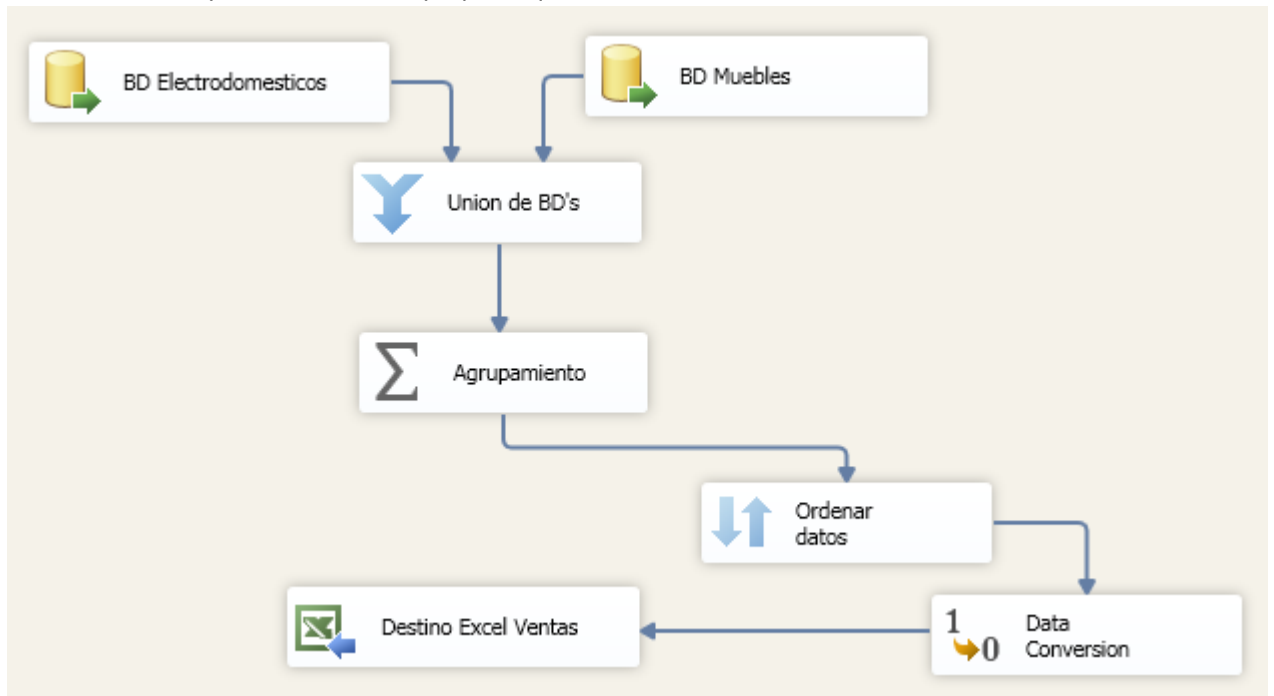
- iii. Hacer clic en OK

- iv. Hacer doble clic en el control Destino Excel Ventas
- v. Hacer clic en Mappings
- vi. Modificar la asignación de columnas así como se muestra a continuación:

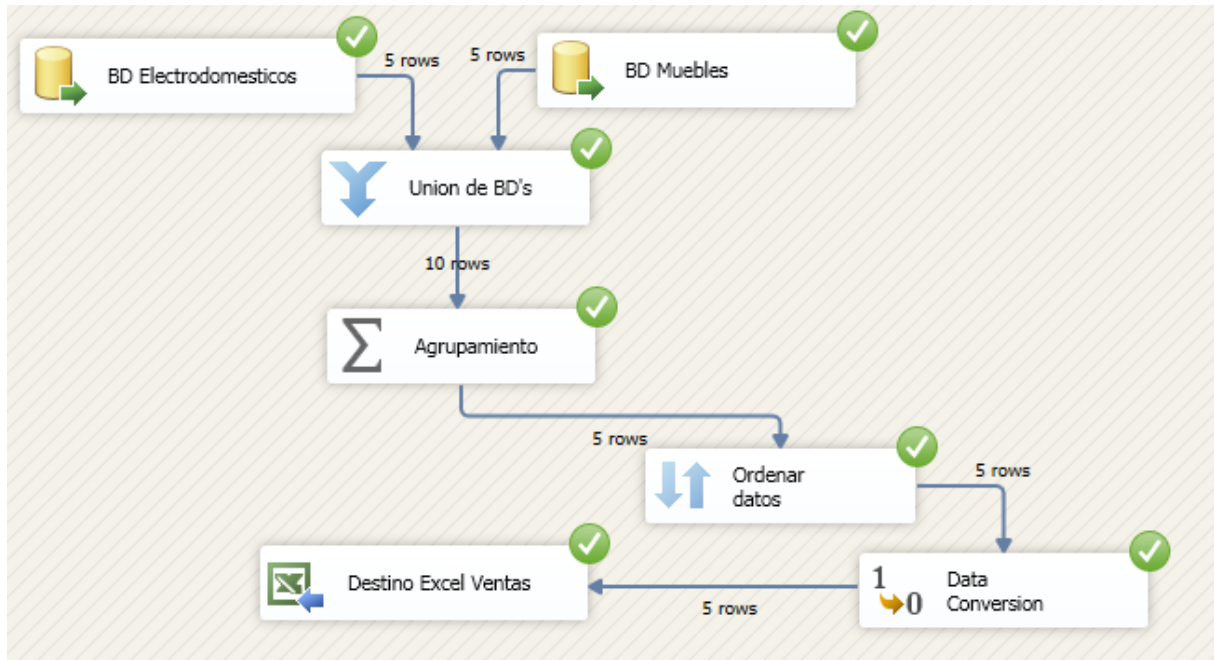


- vii. Hacer clic en OK

17. Con los cambios ya establecidos el paquete queda así:



18. Ejecutar el paquete



Ver que no hay ningún error

19. Abrir el archivo de Excel el cual debe tener los siguientes datos:

Ventas_empleados - Excel

	A	B	C	D	E	F
	CODIGO EMPLEADO	DUI	NIT	APELLIDOS	NOMBRES	VENTAS
1	1	0127859-5	0641140877-193	Hernandez Ramirez	Carlos Roberto	1700.00
2	2	1294587-6	7894120652-147	Perez Hernandez	Maria Leticia	1716.50
3	3	4589785-8	2365110578-145	Carranza	Pablo Jose	1900.40
4	5	6985847-8	4758121280-193	Pastori	Xavier Iñaki	1912.80
5	4	9658745-7	3658090976-145	Rodriguez Arias	Christian Alberto	1915.00

20. Fin de la práctica.

Análisis de resultados

1. Realice modificación al (Ejercicio 3. Crear un proyecto ETL sin el uso del asistente) y agregue un control más de **(OLE DB Source)**

- Crear una nueva base llamada **(db_tecnologia)**
- Crear tabla **(venta_tecnologia)**
- Insertar 5 registros a la tabla

Continuar con los pasos del ejercicio, para generar nuevamente el archivo Excel.

Ejercicios complementarios

Analizar la mejor solución para cargar la información en el modelo tipo estrella del Data Warehouse que se presenta en la imagen 1, cargar como mínimo 10 registros, pueden ser diferentes orígenes de datos o archivos Excel.

De los primeros pasos es crear el modelo de tipo estrella que se presenta en la imagen, para que se pueda realizar la carga de la información por medio de ETL, de los orígenes de datos que ustedes definan.

Los procesos ETL Una vez definido el modelo lógico, comienza la fase de prueba con datos, a través de los procesos de ETL, esto quiere decir que se extraerán datos de diferentes fuentes, para luego integrarlos, filtrarlos y depurarlos.

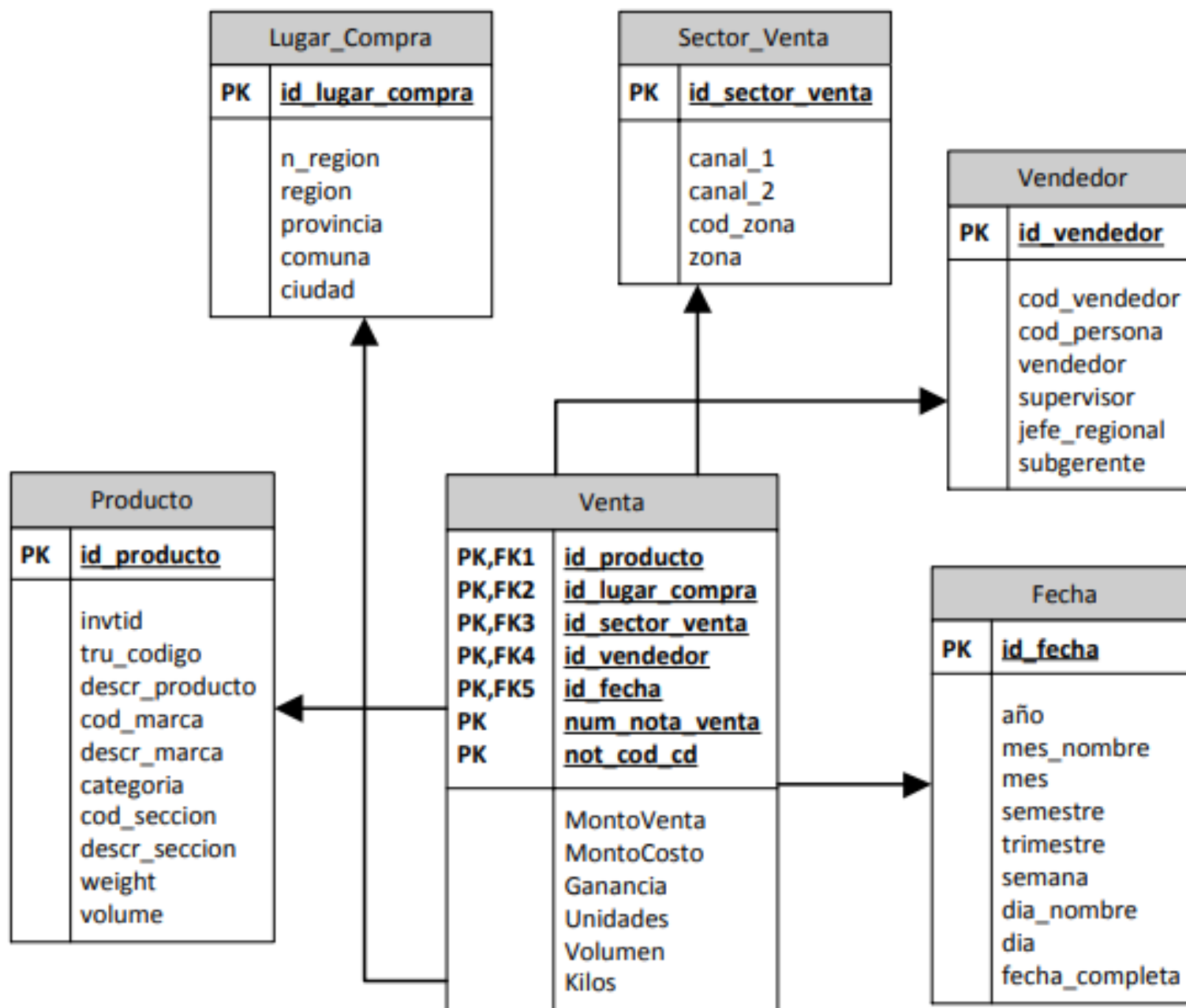


Figura 32: Modelo tipo Estrella del Data Warehouse