

Existen dos formas de tomar decisiones:



Decisiones *lentas*, basadas sólo en la *experiencia*, prediciendo el futuro en base a la *intuición*.



Decisiones *rápidas y oportunas*, tomadas sobre *cimientos firmes*, prediciendo el futuro en base a *información* del pasado.

¿Usted cómo lo haría en su empresa?

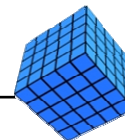


SISTEMAS
DE GESTIÓN II

DOCENTES: Ing. Bigatti, Cristian G.
Est. Grasso, Mónica

COLABORADORES: Alvarez, Ezequiel
Gorza, Luciano
Leala, Guillermo

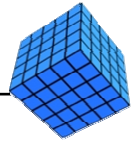




TUTORIAL DATA WAREHOUSE SQL Server: **ETL – Parte II**

Este documento consta de:

- Pasos generales a seguir
- Pasos específicos a seguir



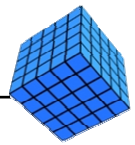
Pasos generales a seguir

Parte I

1. Crear la Base de Datos de la Staging Area.
2. Cargar los datos desde las Fuentes de Datos a la Staging Area.

Parte II

3. Crear la Base de Datos del Data Warehouse.
4. Realizar los procesos de Extracción, Transformación y Carga (ETL).



Pasos específicos a seguir

3. Crear la Base de Datos del Data Warehouse.

a. Ingresaremos al Enterprise Manager y al área de trabajo.

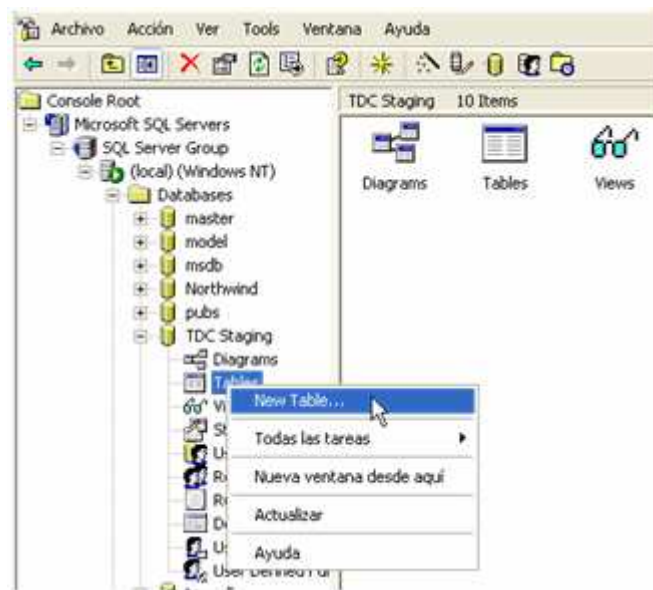
b. Crearemos la Base de datos del Data Warehouse y le daremos un nombre.

Clic con el botón derecho en **Databases** → **New Database....** Escribir el nombre de la base de datos "**TDC DW**". Clic en **Aceptar**.

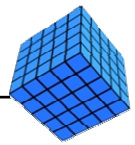
c. Definiremos la estructura de la Base de Datos.

De acuerdo a lo mencionado en la sección **Descripción General del Tutorial**, vamos a crear las tablas que forman parte de la Base de Datos del Data Warehouse.

En **Enterprise Manager**, doble clic en **Databases**, doble clic en **TDC DW**, clic con el botón derecho en **Tables** → **New Table...**



Se abrirá una pantalla que permitirá definir la estructura de la nueva tabla. Esta pantalla funciona de manera similar al modo diseño de Microsoft Access.



	Column Name	Data Type	Length	Allow Nulls
	cod_cliente	bigint	8	
	cod_tiempo	datetime	8	
	cantidad	bigint	8	✓
	litros	float	8	✓

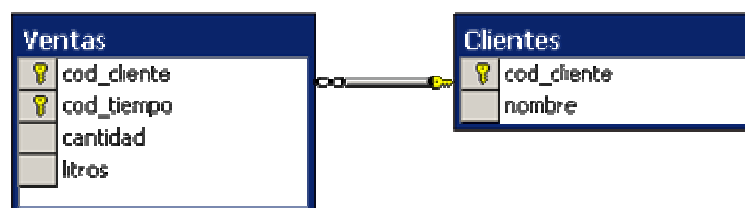
De esta forma crearemos las tablas mencionadas en la sección **Descripción General del Tutorial** del documento **TUTORIAL DATA WAREHOUSE: ETL – Parte I**. Estas tablas son **Cliente** y **Ventas**, y para ellas escribiremos los siguientes atributos y tipos de datos:

Cientes
Cod_cliente (bigint)
Nombre (varchar)

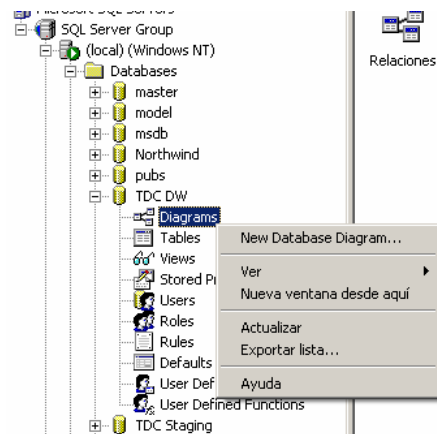
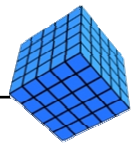
Ventas
Cod_cliente (bigint)
Cod_tiempo (datetime)
Cantidad (bigint)
Litros (float)

Empezaremos con la tabla Clientes y luego con la de Ventas. Al finalizar el diseño de una tabla, clic en el botón **Save**, y escribir el nombre correspondiente. Clic en **OK**.

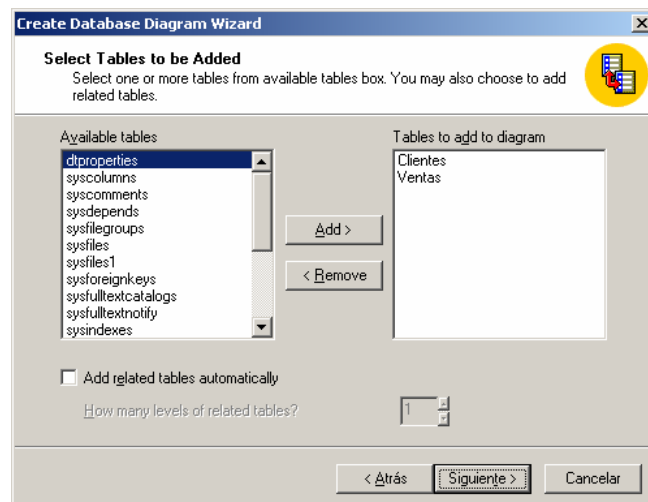
Antes de finalizar con la de Ventas deberemos definir las relaciones entre las tablas:



Para esto, en el **Enterprise Manager**, doble clic en **Databases**, doble clic en **TDC Staging**, clic con el botón derecho en **Diagrams** → **New Database Diagram...**



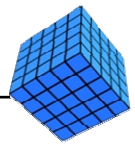
Aparecerá la siguiente ventana, donde seleccionamos las tablas a relacionar:



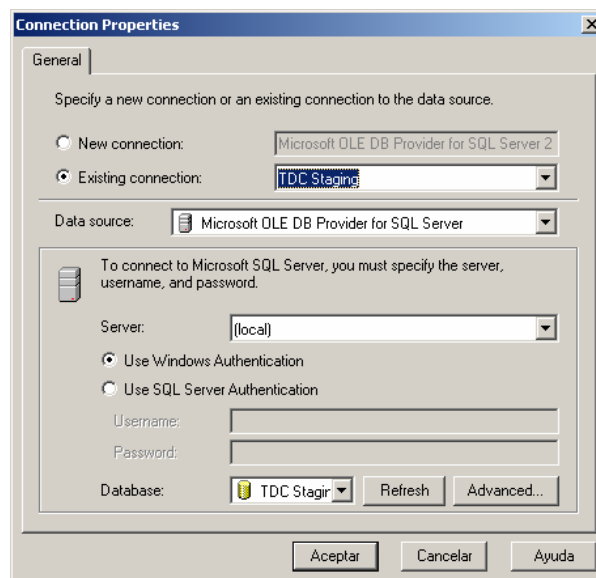
Seleccionamos **Clientes** y **Ventas**, presionamos **Siguiente**, luego finalizar y se abrirá el diagrama con las tablas elegidas y con las relaciones propuestas por SQL Server (siempre que en las tablas estén definidas las claves primarias), las cuales se pueden modificar de ser necesario.

4. Realizar los procesos de Extracción, Transformación y Carga de la Base de Datos del Data Warehouse (ETL).

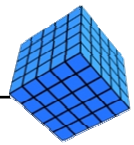
*Los procesos ETL son procesos que leen los registros de las Fuentes de Datos, aplican las transformaciones necesarias para prepararlos y los cargan en el Destino de Datos. Como ejemplo, el proceso **Copy column** mencionado en el documento **TUTORIAL DATA WAREHOUSE: ETL – Parte I** es un proceso ETL.*



- a. Ingresaremos en el DTS Package Designer.
- b. Abriremos el paquete donde realizamos la carga del Staging Area.
- c. Crearemos la transformación para cargar la tabla Clientes del Datawarehouse.
 1. Utilizaremos una conexión para el origen de los datos basada en otra ya existente, que en nuestro caso es el Staging Area, para ello en **DTS Package Designer** → Menú **Connection** → **Microsoft OLE DB Provider for SQL Server....** En el menú desplegable **Existing connection** vamos a seleccionar: **"TDC Staging"**. Clic en **Aceptar**.



2. Crearemos una conexión para el destino de los datos, que en nuestro caso es el Data Warehouse. En **DTS Package Designer** → Menú **Connection** → **Microsoft OLE DB Provider for SQL Server....** En el cuadro de texto **New connection** vamos a escribir un nombre para la conexión: **"TDC DW"**. En el menú desplegable **Database**, seleccionar la base de datos que hemos creado en el punto 3., **TDC DW**. Clic en **Aceptar**.
3. Uniremos a las conexiones fuente con la conexión destino mediante transformaciones simples, de tipo Copy column.



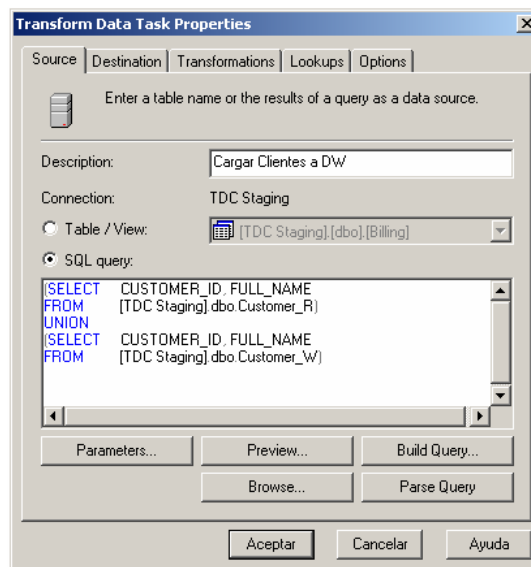
Transformación **TDC Staging** a **TDC DW**: En **DTS Package Designer** → Clic en **TDC Staging**, clic en el botón **Transform Data Task** (representado en la barra de herramientas superior por una flecha ingresando a un engranaje) luego clic en **TDC DW**.

Doble clic sobre la flecha gris que aparece luego de ejecutar el paso anterior. Aparece el cuadro de diálogo Transform Data Task Properties.

En la solapa **Source**, en el cuadro de texto **Description** escribiremos el nombre de la transformación, en nuestro caso **"Cargar Clientes a DW"**.

Luego seleccionamos el option button "SQL query" y escribimos en el cuadro de texto la siguiente consulta:

```
(SELECT  CUSTOMER_ID, FULL_NAME
FROM      [TDC Staging].dbo.Customer_R)
UNION
(SELECT  CUSTOMER_ID, FULL_NAME
FROM      [TDC Staging].dbo.Customer_W)
```

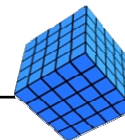


Clic en la solapa **Destination**. Seleccionamos la Tabla Clientes en el combo Table name.

Clic en la solapa **Transformations**, automáticamente se genera el mapeo entre el origen y el destino, a traves de la transformación copy column. Clic en Aceptar.

d. Crearemos la transformación para cargar una nueva tabla en la Staging Area (Staging Ventas).

1. Utilizaremos una conexión para el origen de los datos basada en otra ya existente, que en nuestro caso es el Staging Area, para ello en **DTS**



Package Designer → Menú **Connection** → **Microsoft OLE DB Provider for SQL Server...** En el menú desplegable **Existing connection** vamos a seleccionar: **"TDC Staging"**. Clic en **Aceptar**.

2. Crearemos una conexión para el destino de los datos, que en este caso es el Staging de Ventas. En **DTS Package Designer** → Menú **Connection** → **Microsoft OLE DB Provider for SQL Server...** En el cuadro de texto **New connection** vamos a escribir un nombre para la conexión: **"TDC Staging Ventas"**. En el menú desplegable **Database**, seleccionar la base de datos, **TDC Staging**. Clic en **Aceptar**.

3. Uniremos a las conexiones fuente con la conexión destino.

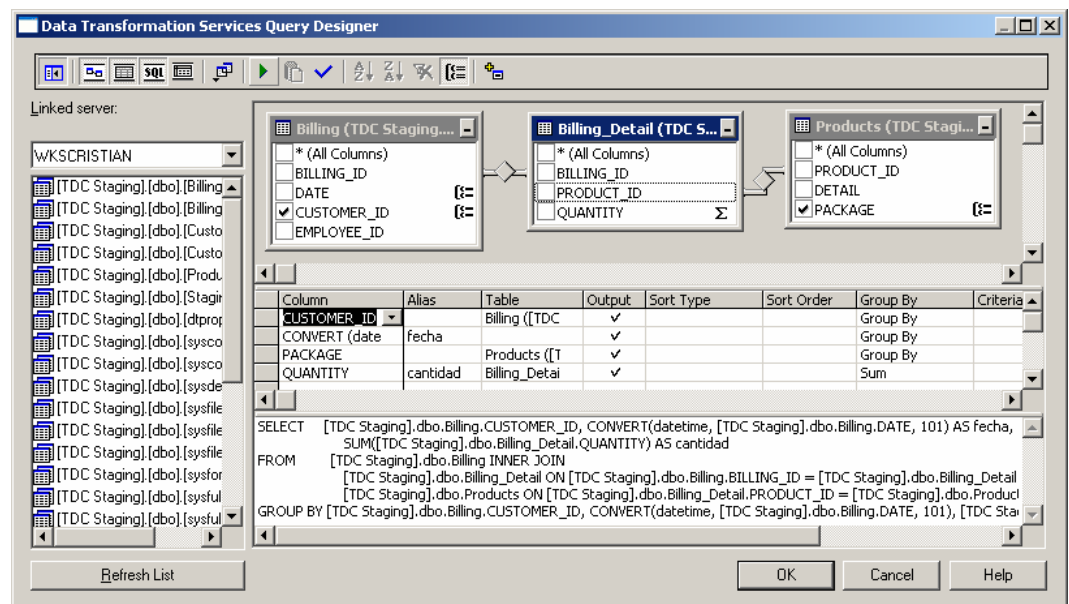
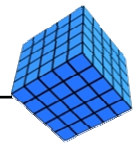
Transformación **TDC Staging** a **TDC Staging Ventas**: En **DTS Package Designer** → Clic en **TDC Staging**, clic en el botón **Transform Data Task**, luego clic en **TDC Staging Ventas**.

Doble clic sobre la flecha gris que aparece luego de ejecutar el paso anterior. Aparece el cuadro de diálogo Transform Data Task Properties.

En la solapa **Source**, en el cuadro de texto **Description** escribiremos el nombre de la transformación, en nuestro caso **"Cargar Staging Ventas"**. Luego seleccionamos el option button "SQL query" y escribimos en el cuadro de texto la siguiente consulta:

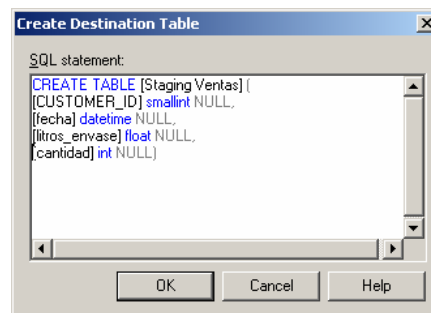
```
SELECT
[TDC Staging].dbo.Billing.CUSTOMER_ID,
CONVERT(datetime, [TDC Staging].dbo.Billing.DATE, 101) AS fecha,
[TDC Staging].dbo.Products.PACKAGE,
SUM([TDC Staging].dbo.Billing_Detail.QUANTITY) AS cantidad
FROM
[TDC Staging].dbo.Billing INNER JOIN
[TDC Staging].dbo.Billing_Detail ON
[TDC Staging].dbo.Billing.BILLING_ID =
[TDC Staging].dbo.Billing_Detail.BILLING_ID
INNER JOIN [TDC Staging].dbo.Products ON
[TDC Staging].dbo.Billing_Detail.PRODUCT_ID =
[TDC Staging].dbo.Products.PRODUCT_ID
GROUP BY
[TDC Staging].dbo.Billing.CUSTOMER_ID,
CONVERT(datetime,[TDC Staging].dbo.Billing.DATE, 101),
[TDC Staging].dbo.Products.PACKAGE
```

Opcionalmente se podría crear la consulta mediante el boton **Build Query...**, el cual permitirá crear en forma gráfica la sentencia SQL:

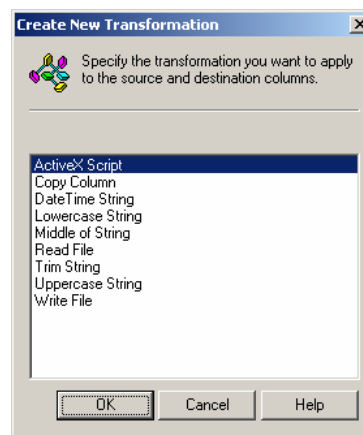
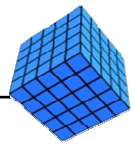


Clic en la solapa **Destination**. Hacemos clic en **Create....** En el cuadro de texto ingresamos:

```
CREATE TABLE [Staging Ventas] (
    [CUSTOMER_ID] smallint NULL,
    [fecha] datetime NULL,
    [litros_envase] float NULL,
    [cantidad] int NULL)
```



Clic en **OK**. Clic en la solapa **Transformations**. Borramos la transformación PACKAGE a litros_envase, presionando en la flecha y luego presionando **Delete**. Luego Seleccionamos PACKAGE en el origen y litros_envase en el destino, después presionamos **New** para crear un nuevo tipo de transformación, en este caso "ActiveX Script". Aparecerá la siguiente ventana:



Seleccionamos ActiveX Script y presionamos **OK**. Luego en la solapa **General** presionamos **Properties**. En el cuadro de texto reemplazar lo escrito por el siguiente script:

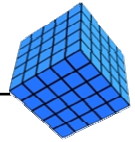
```
Function Main()  
    Select Case DTSSource("PACKAGE")  
        Case "1 Liter" DTSDestination("litros_envase")=1  
        Case "2 Liter" DTSDestination("litros_envase")=2  
        Case "330 cm3 can" DTSDestination("litros_envase")=0.33  
        Case "500 cm3 can" DTSDestination("litros_envase")=0.5  
        Case "670 cm3" DTSDestination("litros_envase")=0.67  
    End Select  
    Main = DTSTransformstat_InsertQuery  
End Function
```

Presionamos **OK, Aceptar, Aceptar**.

e. Crearemos la transformación para cargar la tabla Ventas del Data Warehouse.

1. Utilizaremos una conexión para el origen de los datos basada en otra ya existente, que en nuestro caso es el TDC Staging Ventas, para ello en **DTS Package Designer** → Menú **Connection** → **Microsoft OLE DB Provider for SQL Server....** En el menú desplegable **Existing connection** vamos a seleccionar: **"TDC Staging Ventas"**. Clic en **Aceptar**.

2. Utilizaremos una conexión para el destino de los datos basada en otra ya existente, que en nuestro caso es el TDC DW, para ello en **DTS Package Designer** → Menú **Connection** → **Microsoft OLE DB Provider for SQL Server....** En el menú desplegable **Existing connection** vamos a seleccionar: **"TDC DW"**. Clic en **Aceptar**.



3. Uniremos a las conexiones fuente con la conexión destino.

Transformación **TDC Staging Ventas** a **TDC DW**: En **DTS Package Designer** → Clic en **TDC Staging Ventas**, clic en el botón **Transform Data Task**, luego clic en **TDC DW**.

Doble clic sobre la flecha gris que aparece luego de ejecutar el paso anterior. Aparece el cuadro de diálogo Transform Data Task Properties.

En la solapa **Source**, en el cuadro de texto **Description** escribiremos el nombre de la transformación, en nuestro caso "**Carga Ventas**". Luego seleccionamos el option button "SQL query" y escribimos en el cuadro de texto la siguiente consulta:

```
SELECT
CUSTOMER_ID, fecha
SUM(litros_envase * cantidad) AS litros,
SUM(cantidad) AS cantidad
FROM
[TDC Staging].dbo.[Staging Ventas]
GROUP BY
CUSTOMER_ID, fecha
```

Clic en la solapa **Destination**. Seleccionamos la Tabla Ventas en el combo Table name.

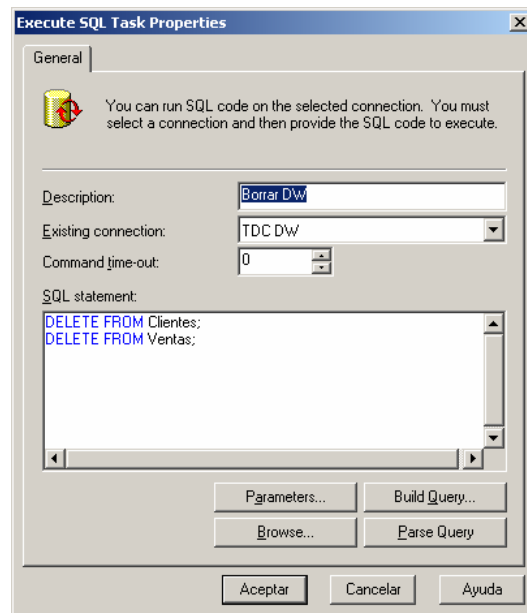
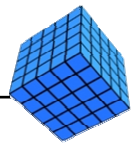
Clic en la solapa **Transformations**, automáticamente se genera el mapeo entre el origen y el destino, a través de la transformación copy column. Clic en **Aceptar**.

- f. Ahora vamos a crear un **Execute SQL Task**, o tarea de ejecución SQL. Vamos a utilizarla para borrar los datos que el proceso almacenará en la Base de Datos **TDC DW**, de tal forma que cada vez que este proceso se corra, se haga una actualización completa (eliminación y reemplazo de los datos) de dicha Base de Datos, a partir de las Fuentes de Datos.

En **DTS Package Designer**, menú **Task** → **Execute SQL Task....** En el cuadro de texto **Description**, escribimos "**Borrar DW**". En el menú desplegable **Existing connection** seleccionamos **TDC DW**. En el cuadro de texto **SQL Statement** escribimos la siguiente sentencia SQL:

```
DELETE FROM Clientes;
DELETE FROM Ventas;
```

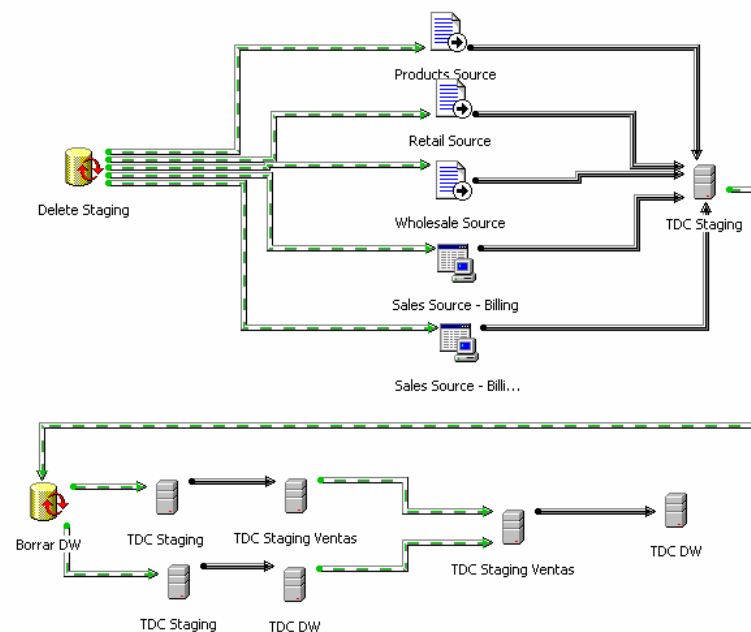
Clic en **Parse Query** para verificar la sintaxis de la sentencia. Si es correcta se mostrará un mensaje de éxito. Clic en **OK**, clic en Aceptar.

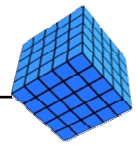


- g. Actualizar el SQL Task Delete staging con la ultima sentencia:

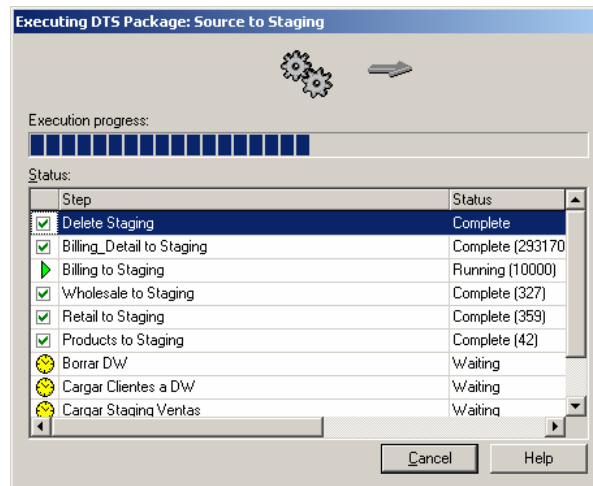
```
DELETE FROM Products;
DELETE FROM Customer_R;
DELETE FROM Customer_W;
DELETE FROM Billing;
DELETE FROM Billing_Detail;
DELETE FROM Staging Ventas;
```

- h. Por último, definiremos la secuencia de tareas mediante condiciones de éxito, de tal forma que quede como indica la siguiente imagen:

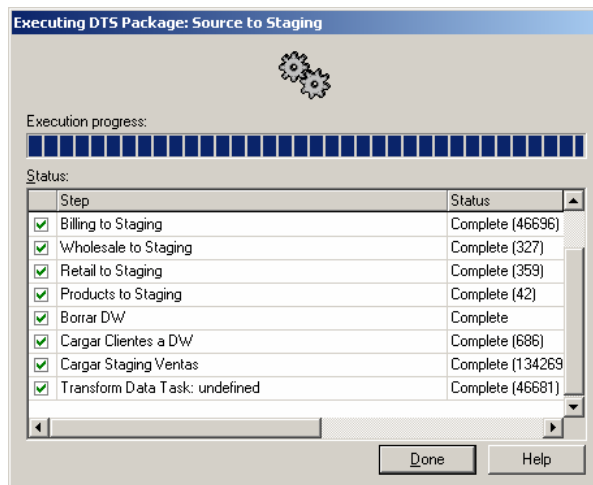




- i. Correremos el proceso. Clic en el botón **Play** (representado en la barra de herramientas superior por un triángulo verde apuntando hacia la derecha).



Si el proceso está correctamente armado, cada una de sus partes se ejecutará en forma satisfactoria. Clic en **Aceptar**, clic en **Done**



En este momento tenemos las tablas de la Base de Datos del Data Warehouse cargadas con los datos operacionales.