

Universidad Rafael Landívar

Facultad de Ingeniería

Análisis de Datos I



Manual Técnico Proyecto - Auto Partes

Derly Rodas 1177318

Pablo Muralles 1113818

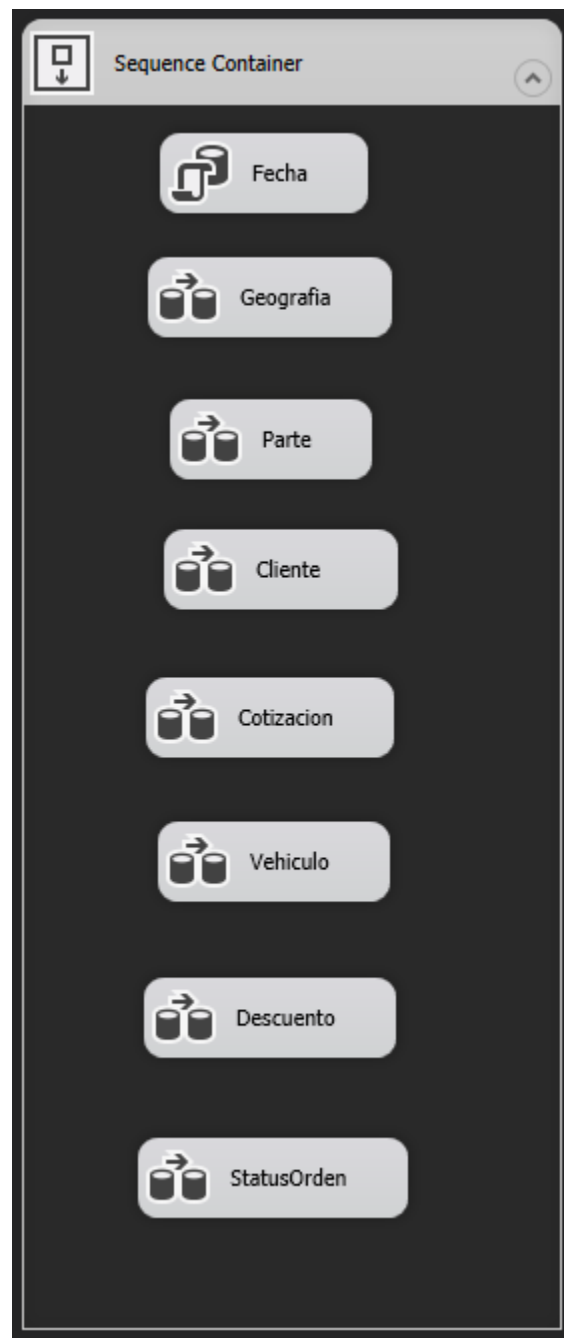
Roberto Solares 1173318

Brenner Hernandez 1023718

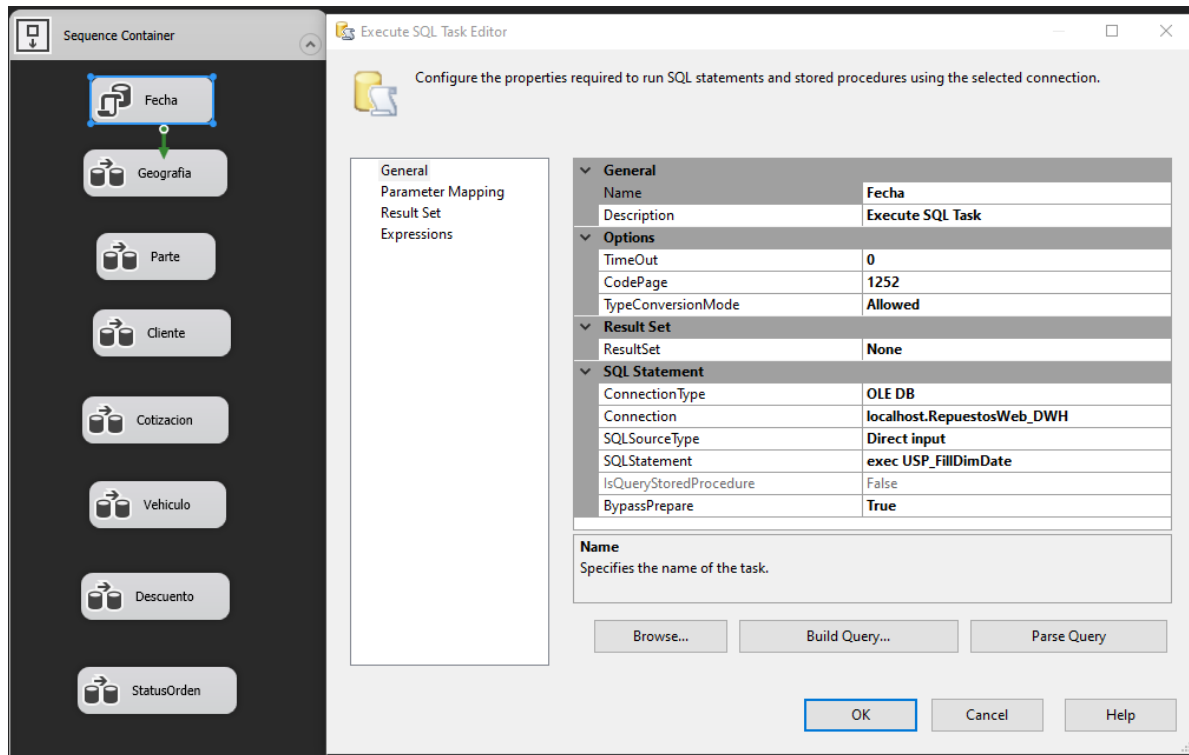
Guatemala, 25 de octubre de 2021

ETL:

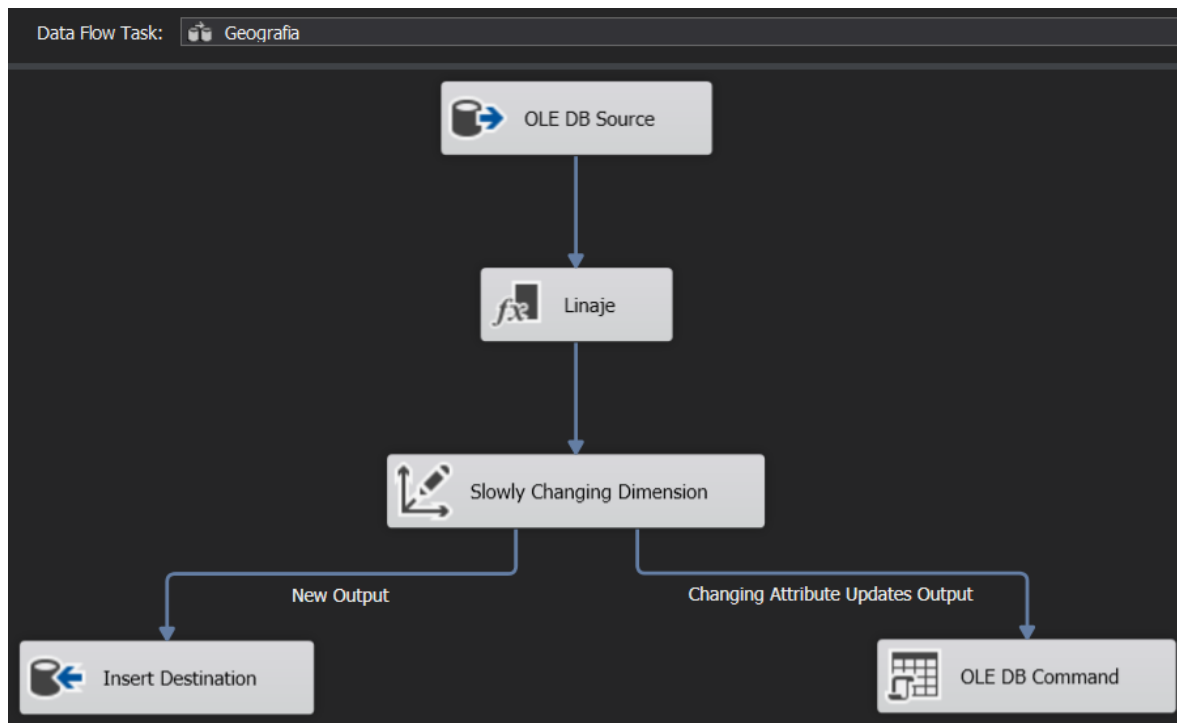
Package Dimension:



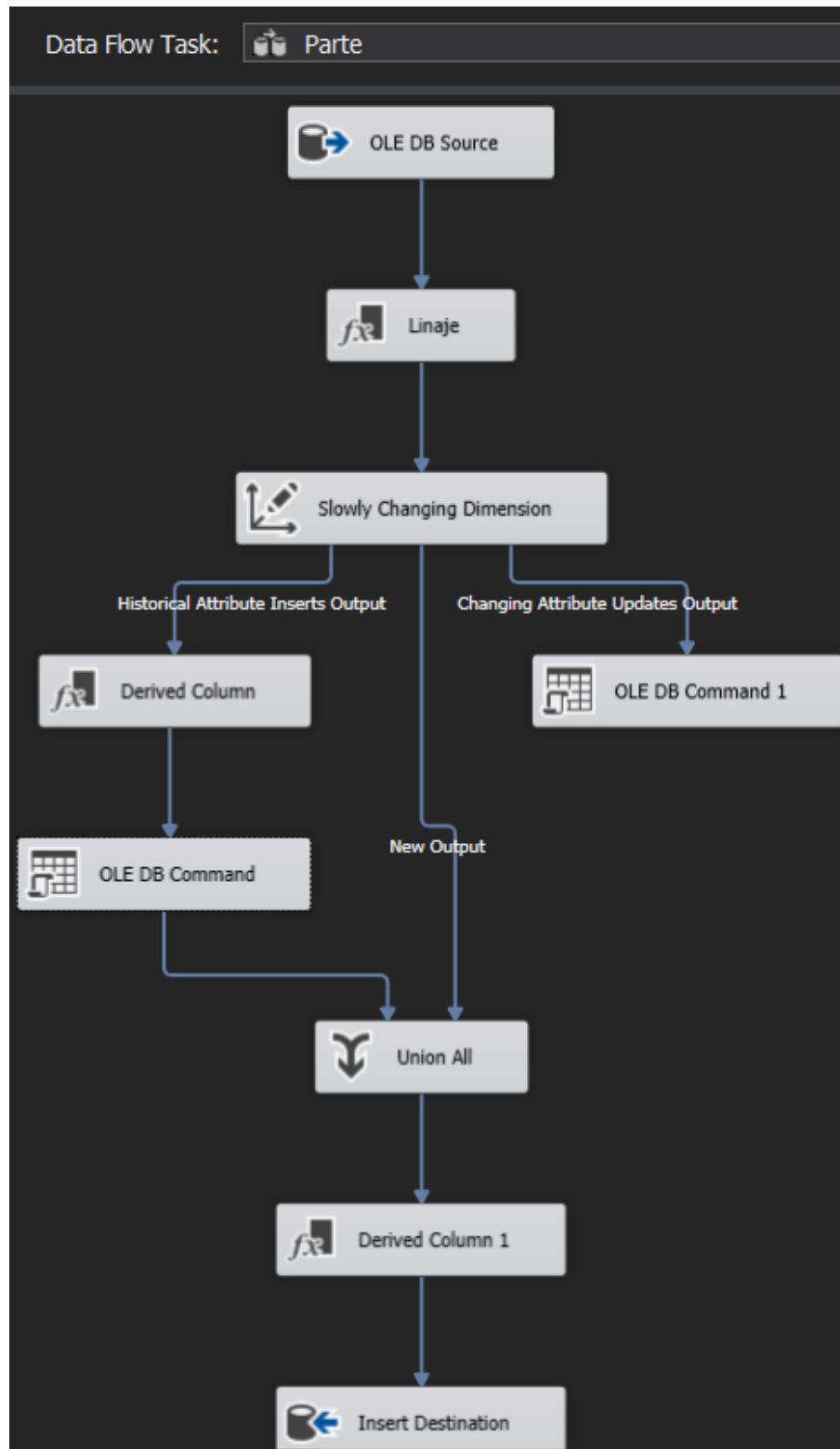
Fecha: Se utilizó un “Execute SQL Task” para ejecutar el procedimiento “USP_FillDimDate” creado con anterioridad



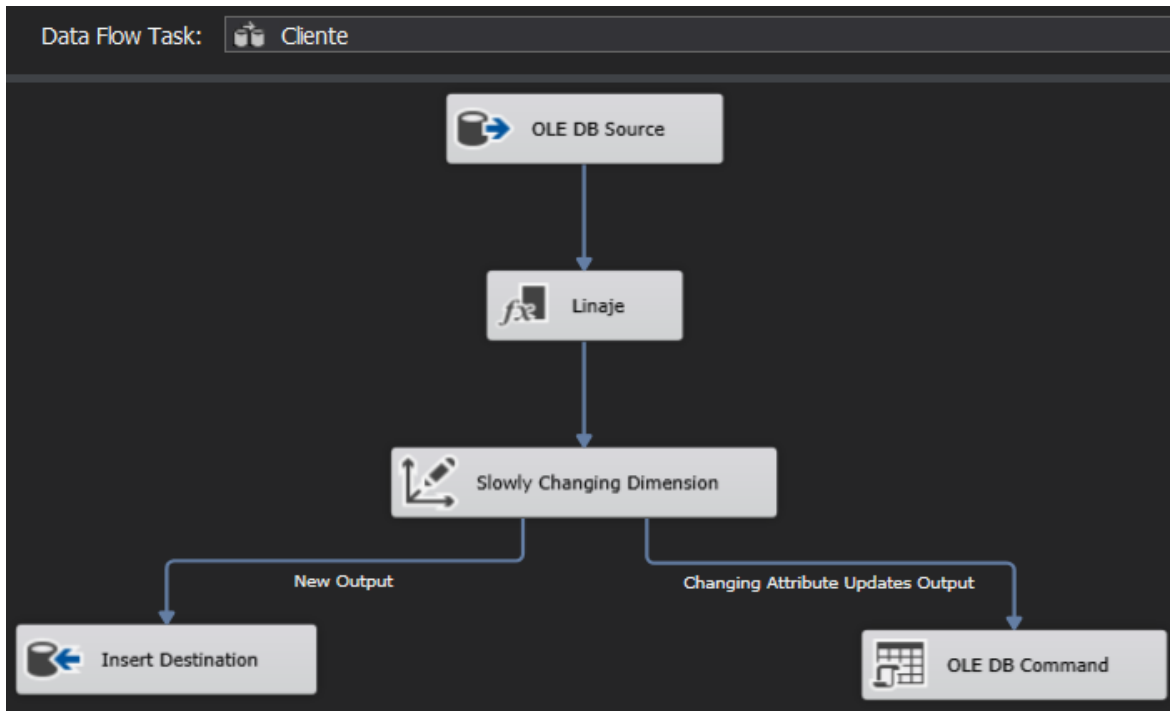
Geografia: Proceso que agrupa la información de las tablas Ciudad, País y Región para luego hacer un SCD donde al código postal se le asigna un tipo fijo (Tipo 0).



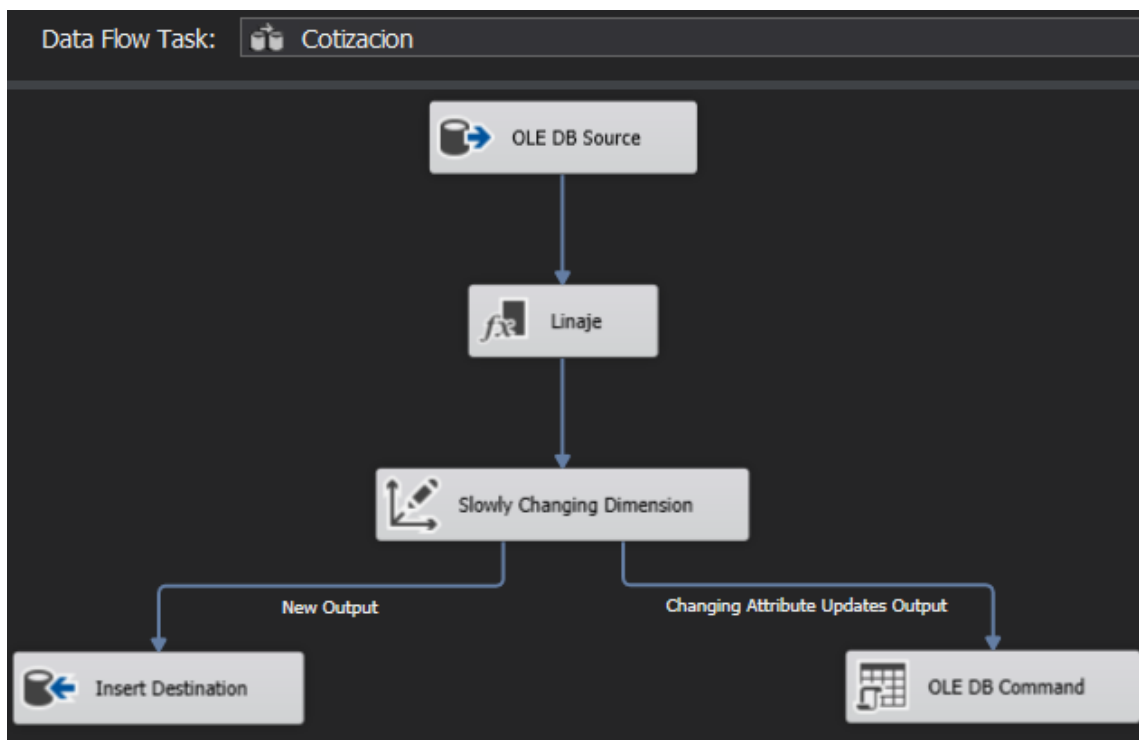
Parte: Para este proceso se utilizó un data flow task que consultaba la información de las tablas Parte, Categoría y Línea para después utilizarla en un proceso SCD donde el precio es de tipo fijo (Tipo 0).



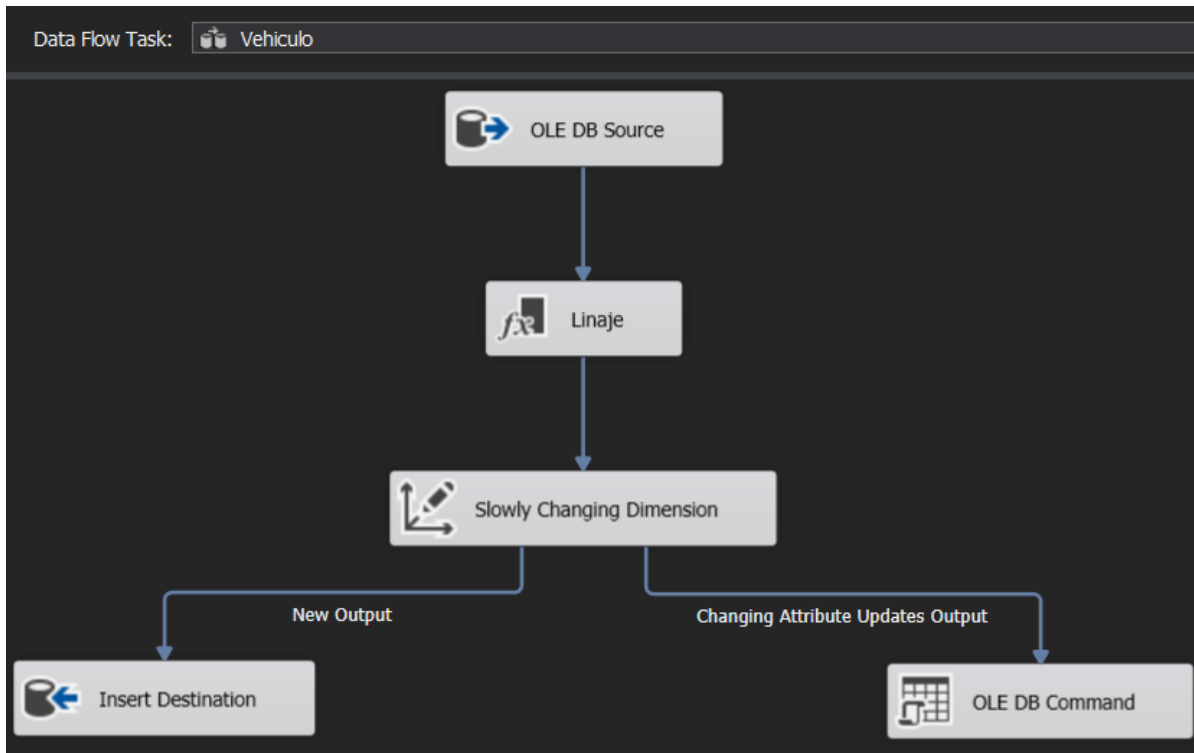
Ciente: Este proceso es un simple data flow task que obtiene la información de la tabla Ciente y utiliza un SCD con todos los datos con tipo cambiante (Tipo 1).



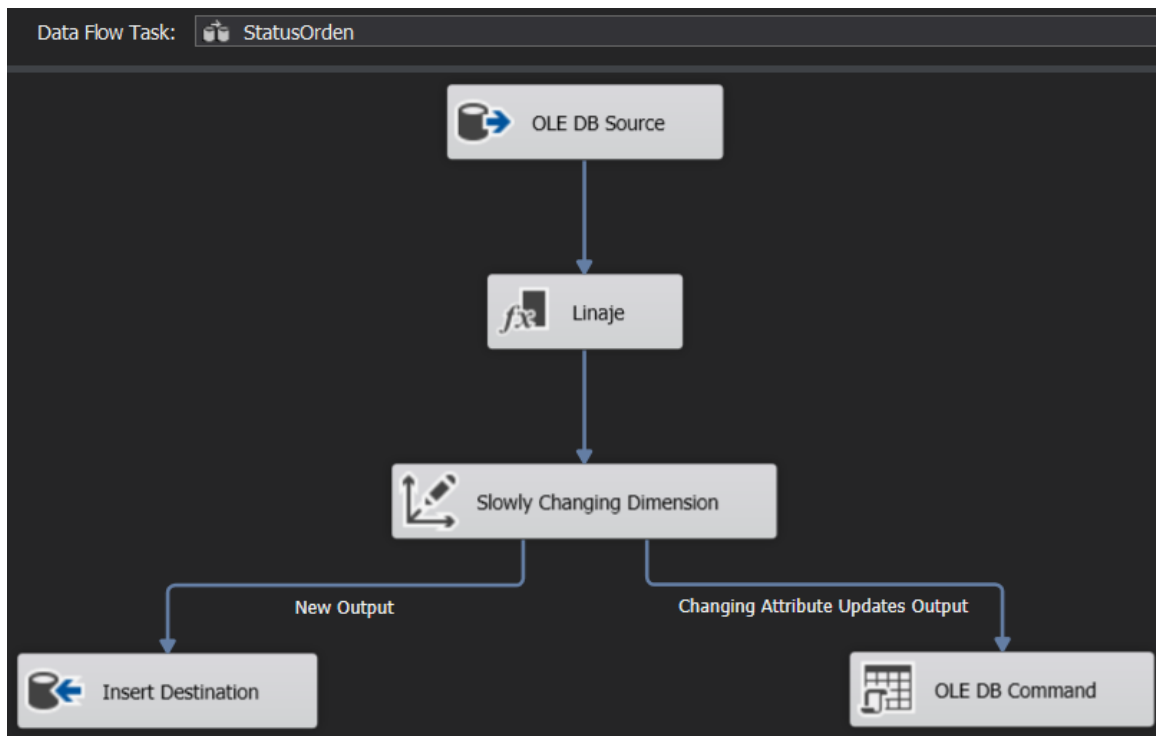
Cotización: Para este proceso se utilizó un data flow task que consultaba la información de las tablas Parte, Categoría y Línea para después utilizarla en un proceso SCD donde el precio es de tipo histórico (Tipo 0).



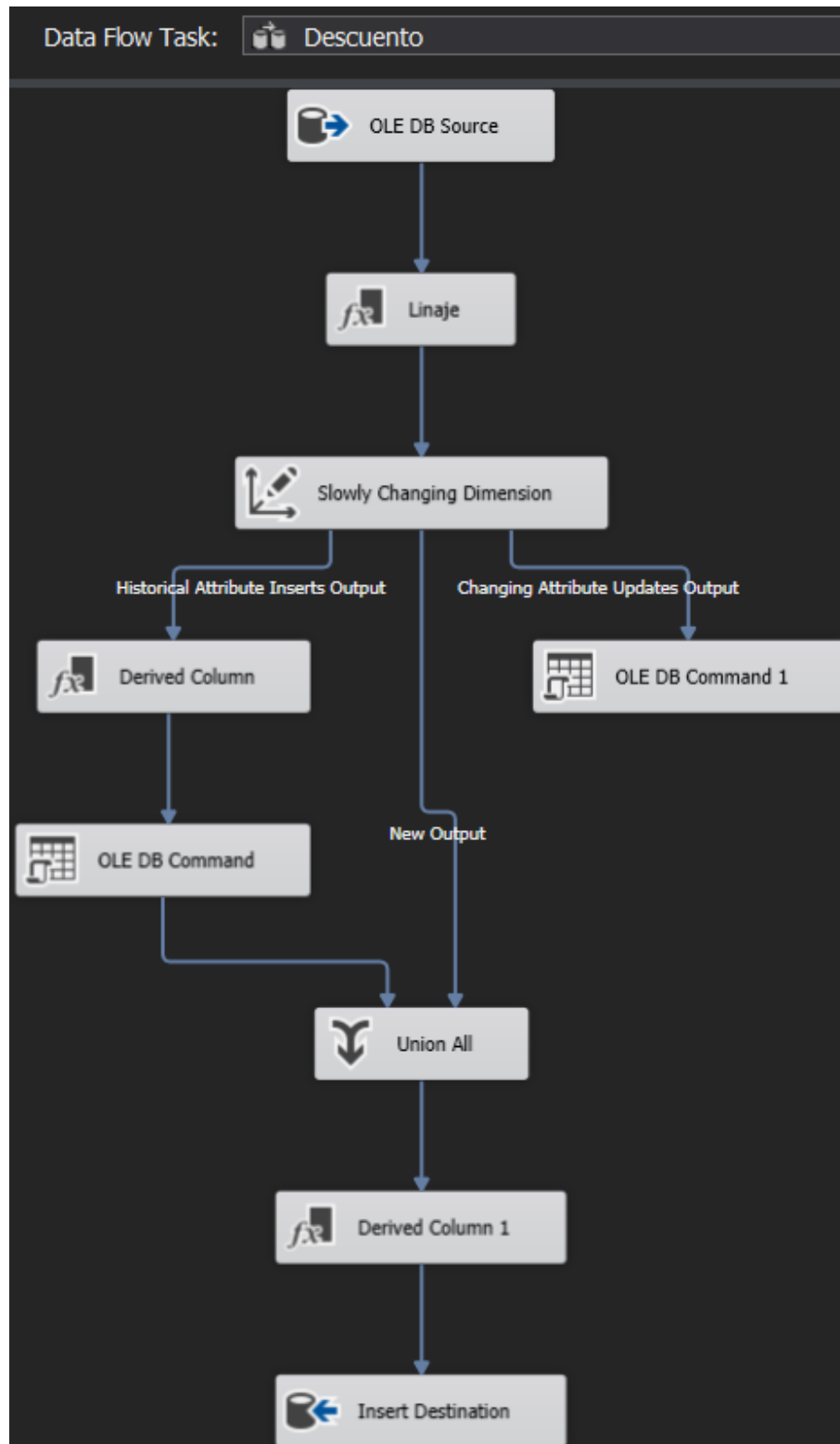
Vehículo: Proceso que agrupa la información de las tabla Vehículo para luego hacer un SCD donde al VIN se le asigna un tipo histórico (Tipo 2).



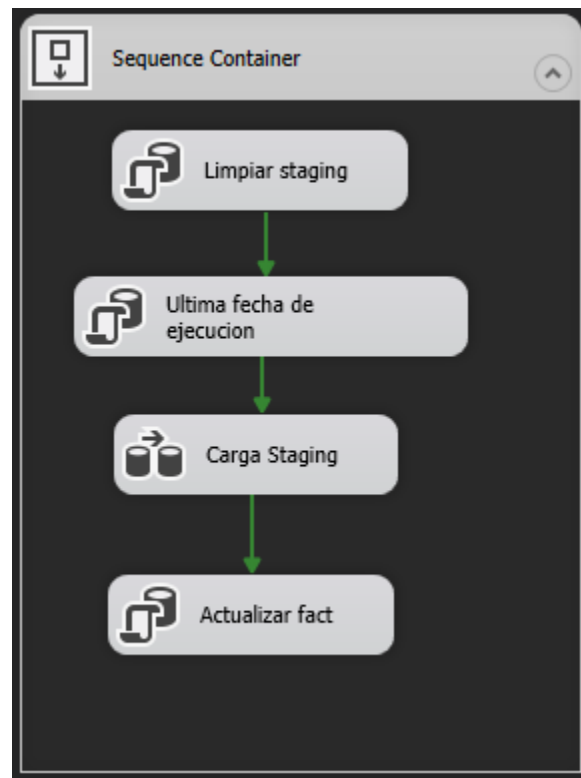
StatusOrden: Este proceso es un simple data flow task que obtiene la información de la tabla StatusOrden y utiliza un SCD con todos los datos con tipo cambiante (Tipo 1).



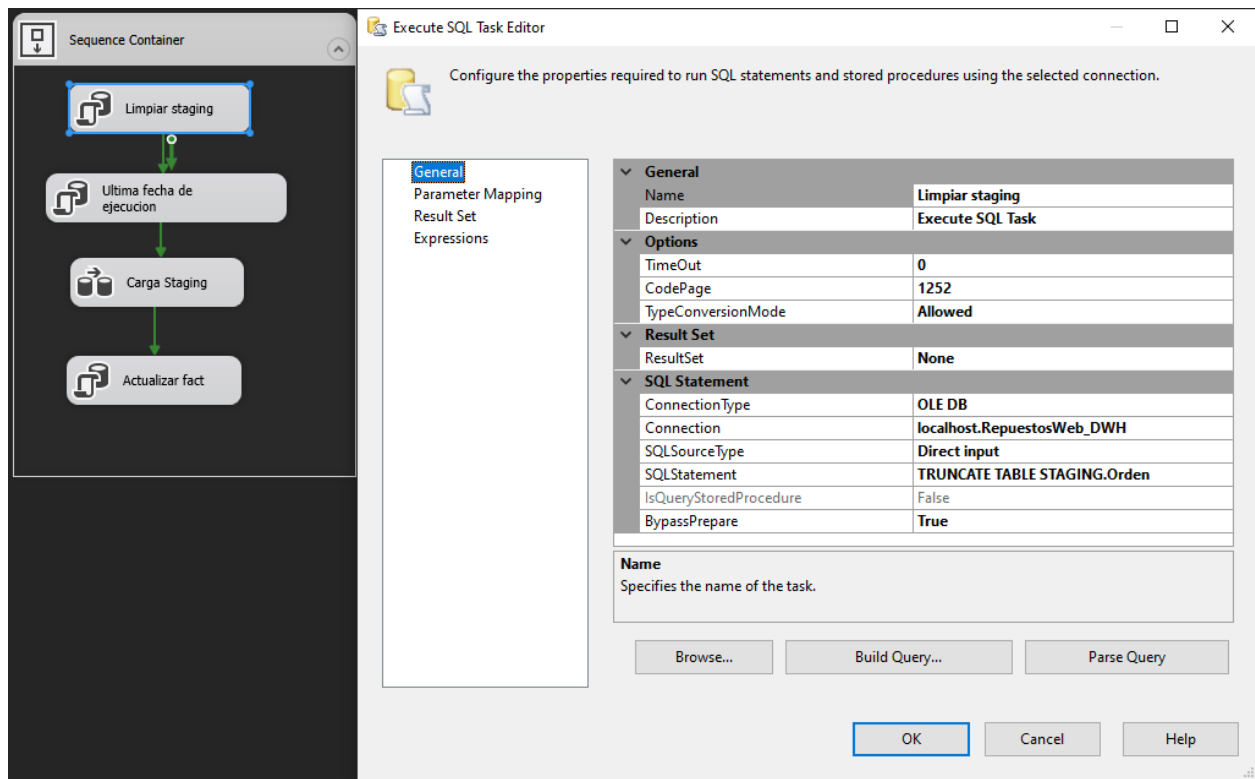
Descuento: Este proceso es un simple data flow task que obtiene la información de la tabla Descuento y utiliza un SCD donde al campo del porcentaje se asigna un valor de tipo histórico (Tipo 2).



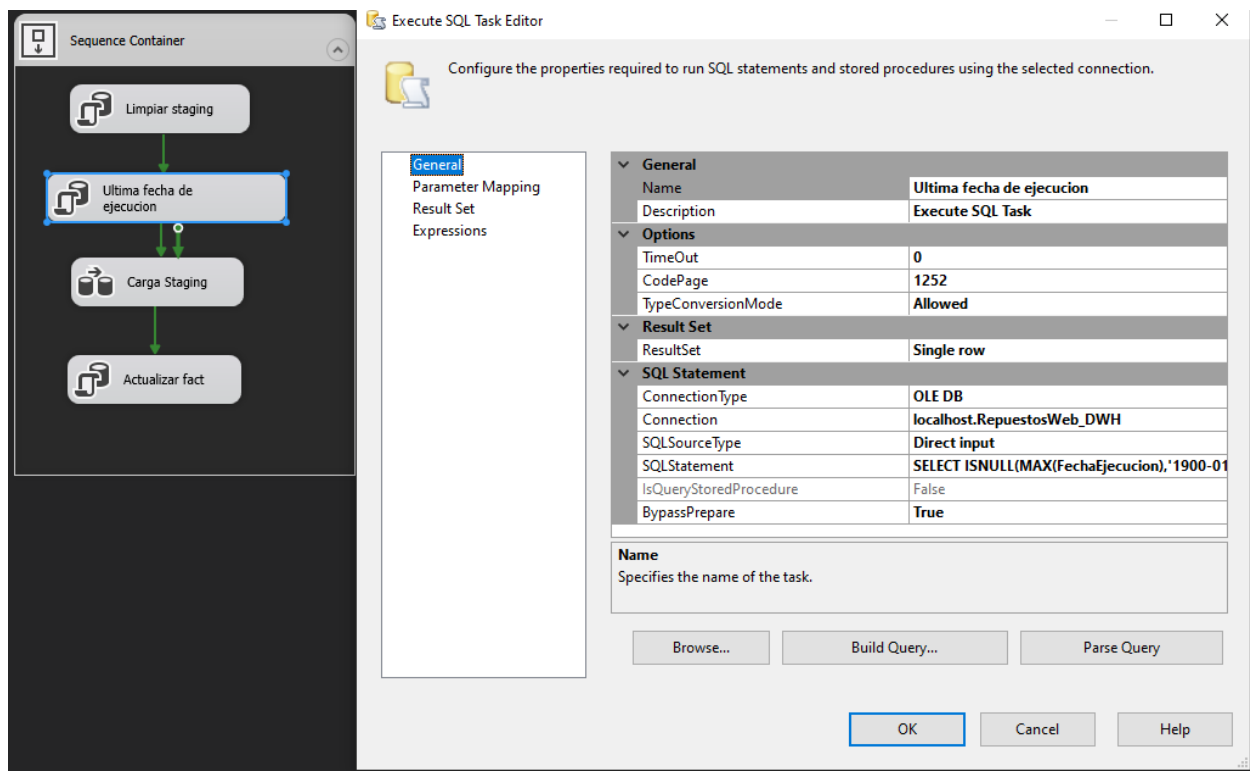
Package Fact:



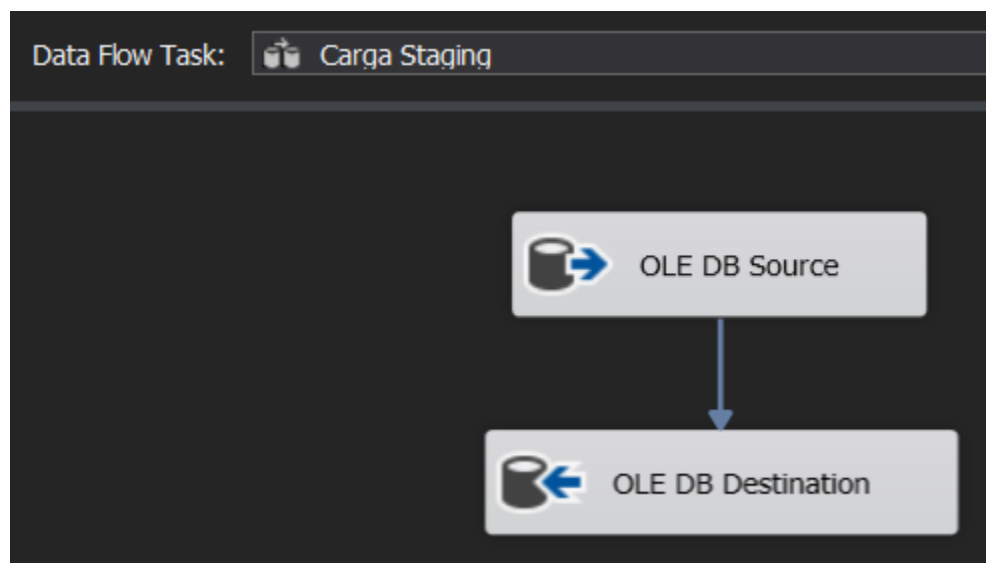
Limpiar staging: Ejecuta un TRUNCATE a la tabla de STAGING.Orden



Última fecha de ejecución: Se realiza un SELECT hacia la FactLog para obtener la última fecha y almacenarla en la variable vFechaEjecución. Que servirá para la carga de data a Staging.



Carga Staging: Este proceso carga los datos que se encuentran en las tablas Orden y DetalleOrden, y los guarda en la tabla de staging del datawarehouse. Para este proceso se utiliza la variable anteriormente descrita para cargar solo los datos nuevos a la tabla de staging.



Data Flow Task: Carga Staging

```
graph TD; Source[OLE DB Source] --> Destination[OLE DB Destination];
```

OLE DB Source Editor

Configure the properties used by a data flow to obtain data from any OLE DB provider.

Connection Manager
Columns
Error Output

OLE DB connection manager:
localhost.RepuestosWeb

Data access mode:
SQL command

SQL command text:

```
SELECT o.ID_Orden,
o.ID_Cliente,
o.ID_Ciudad,
do.ID_DetalleOrden,
do.ID_Parte,
do.ID_Descuento,
do.VehiculoID,
o.ID_StatusOrden,
o.Total_Orden,
o.Fecha_Orden,
o.NumeroOrden,
do.Cantidad
FROM DBO.Orden o
INNER JOIN DBO.Detalle_Orden do ON(o.ID_Orden =
```

Preview...

Connection Managers

localhost.RepuestosWeb	localhost.RepuestosWeb_DWH
------------------------	----------------------------

Data Flow Task: Carga Staging

```
graph TD; Source[OLE DB Source] --> Destination[OLE DB Destination];
```

OLE DB Destination Editor

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

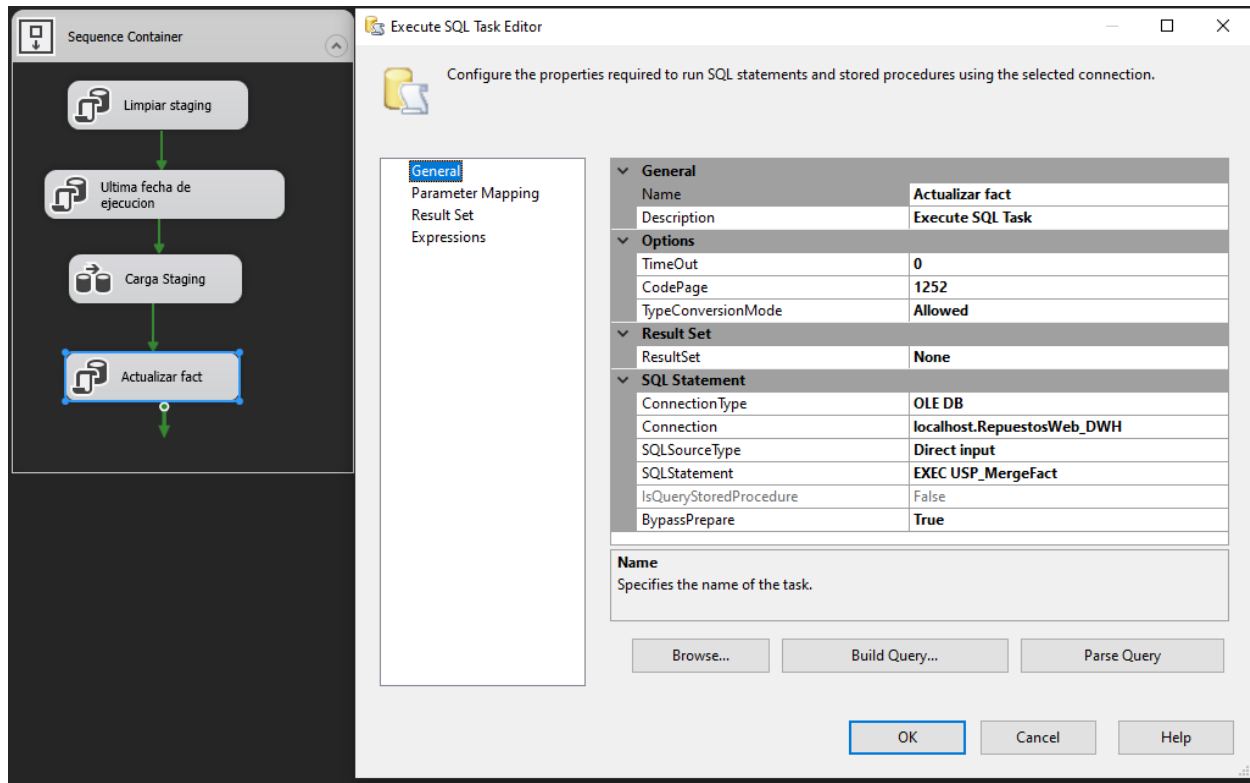
Connection Manager
Mappings
Error Output

OLE DB connection manager:
localhost.RepuestosWeb_DWH

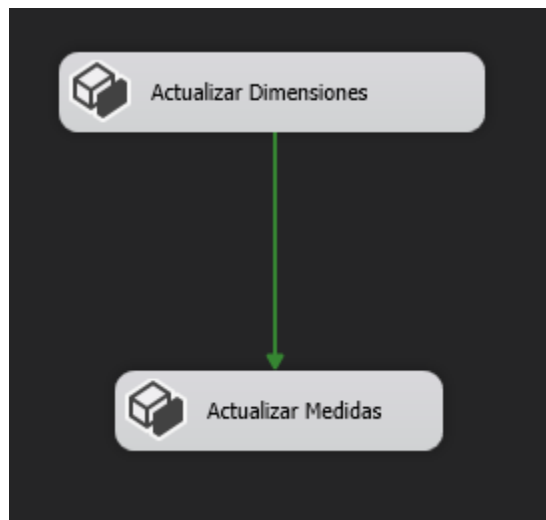
Data access mode:
Table or view

Name of the table or the view:
[staging].[Orden]

Actualizar fact: Se utilizó un “Execute SQL Task” para ejecutar el procedimiento “USP_MergeFact” creado con anterioridad.



Package Cubo:



Actualizar Medidas y Dimensiones: Se encarga de actualizar los datos del cubo a partir del DWH.

SSAS:

Data Source: Se conecta a la base de datos de Repuestos Web DWH para obtener los datos para el cubo.

Data Source Designer

General | Impersonation Information

Data source name: Repuestos Web DWH

Provider:

Connection string: Provider=SQLNCLI11.1;Data Source=localhost;Integrated Security=SSPI; Edit...

Data source references

☐ Maintain a reference to another object in the solution

Create a data source based on an existing data source

Isolation: ReadCommitted

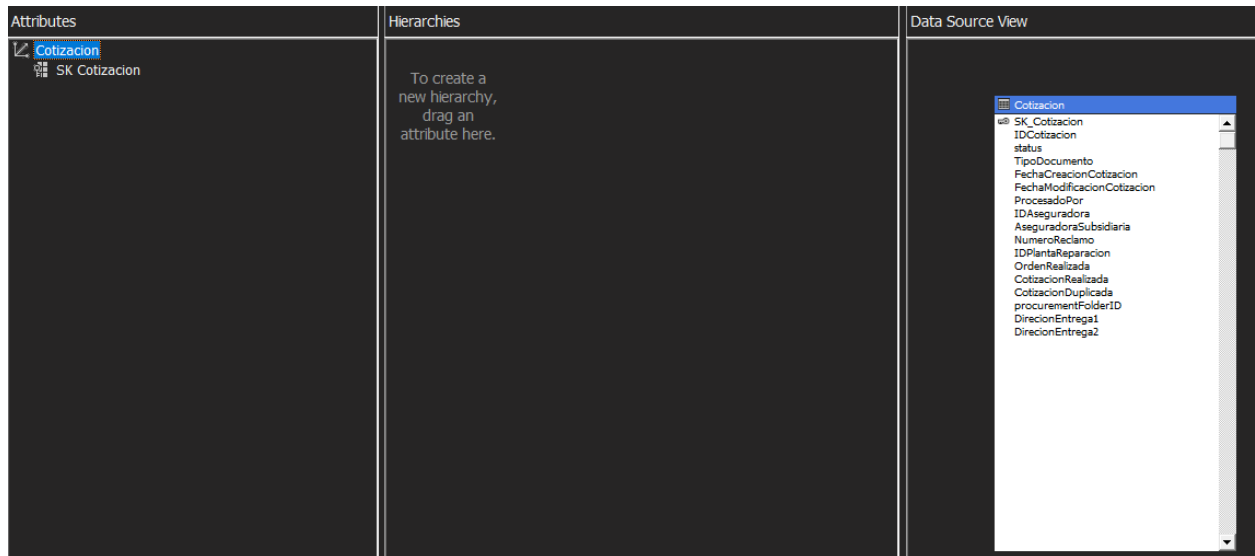
Query timeout (in seconds): 0

Maximum number of connections: 10

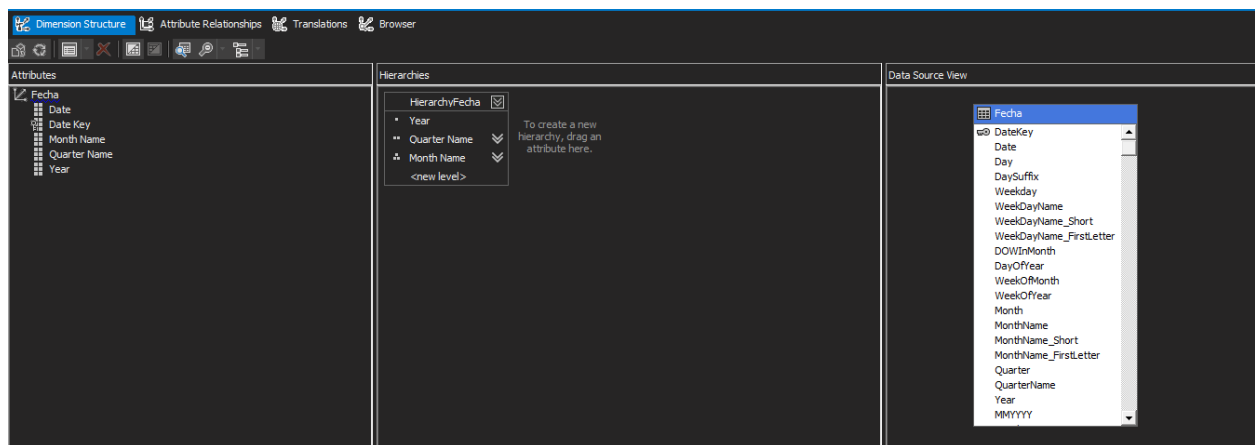
Data source description:

OK Cancel Help

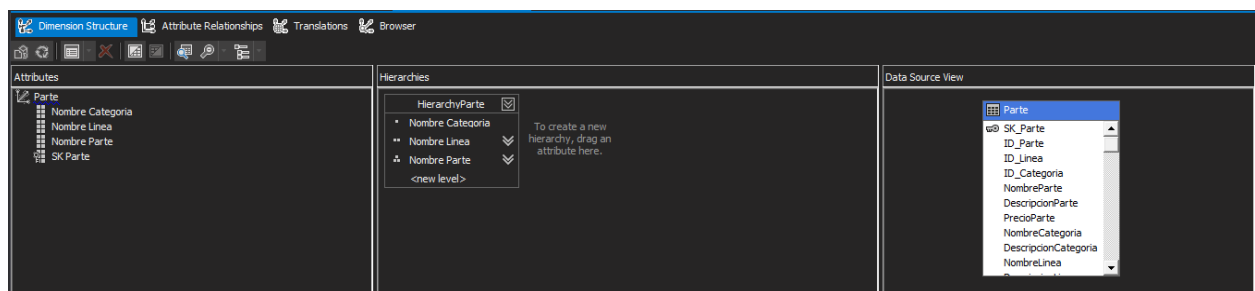
Dimensión Cotización: Esta dimensión se obtiene de la dimensión que se encuentra en el datawarehouse.



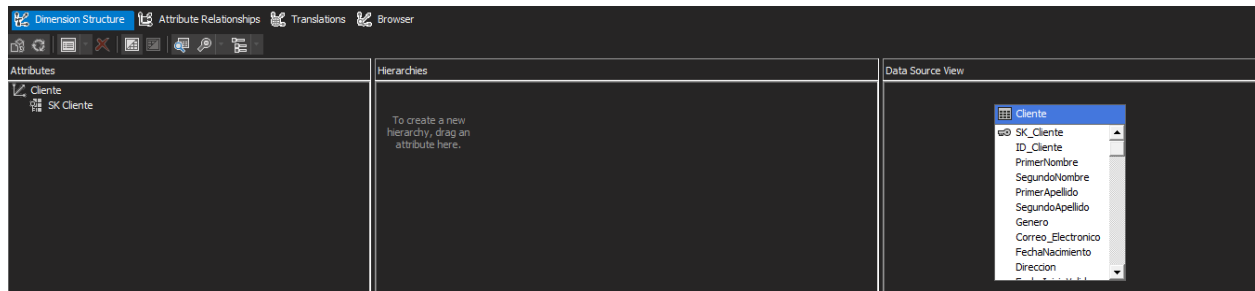
Dimensión Fecha: Esta dimensión se le realizó una jerarquía con la estructura Mes/Cuarto/Año.



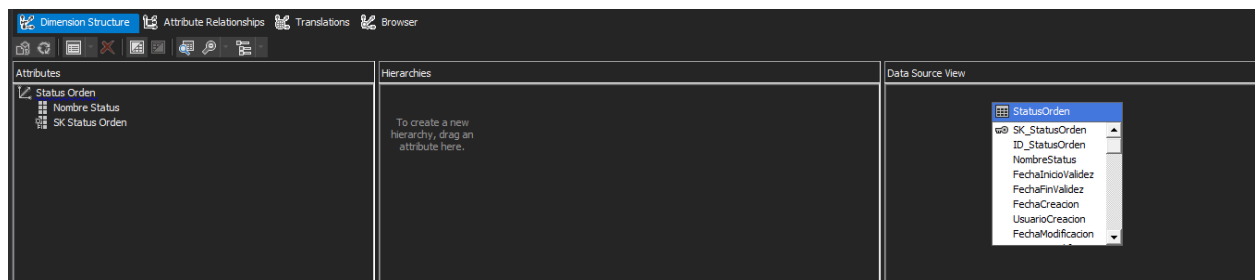
Dimensión Parte: Esta dimensión se le realizó una jerarquía con la estructura Nombre Categoría/Nombre Línea/Nombre Parte.



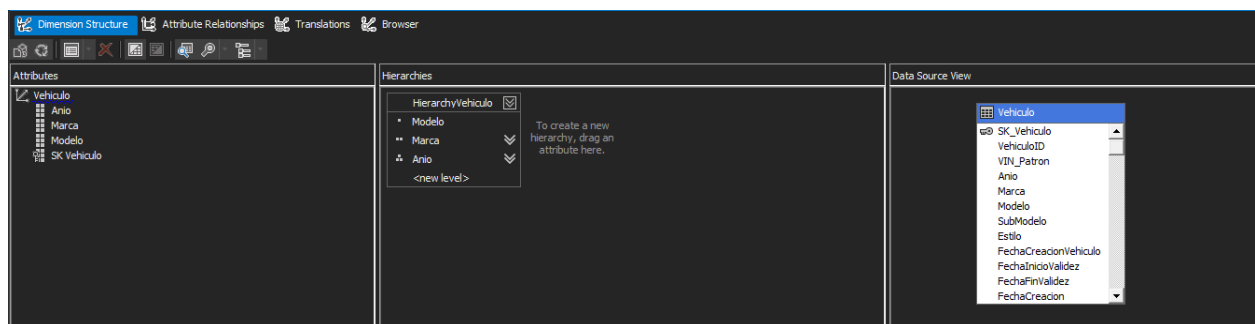
Dimensión Cliente: Esta dimensión se obtiene de la dimensión que se encuentra en el datawarehouse.



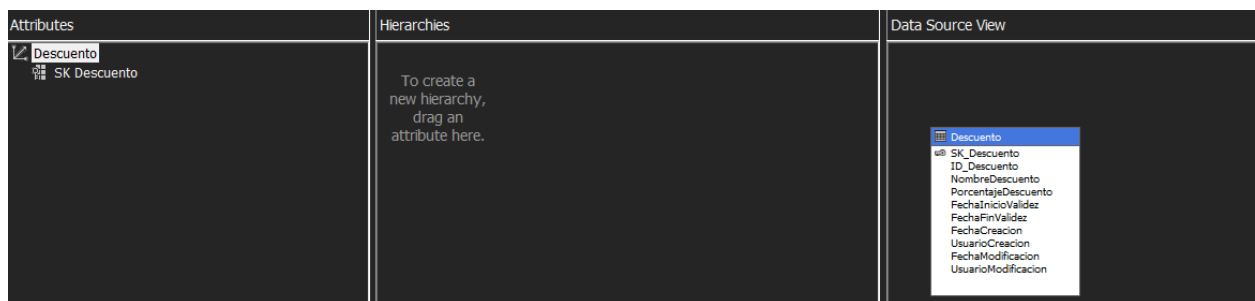
Dimensión Status Orden: Esta dimensión se obtiene de la dimensión que se encuentra en el datawarehouse.



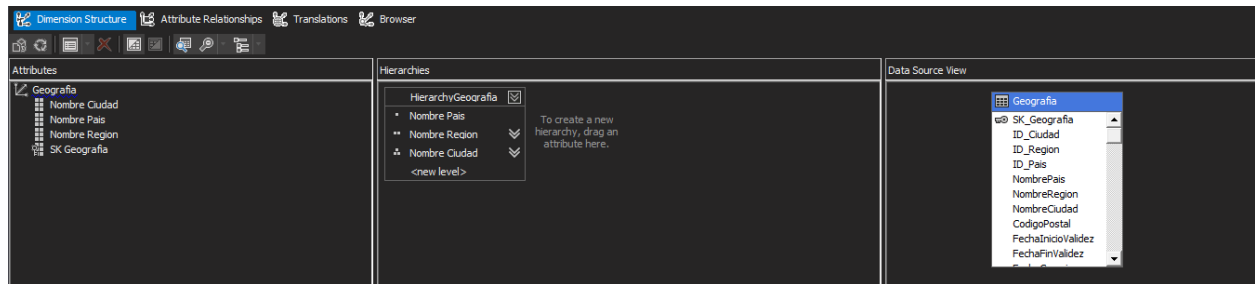
Dimensión Vehículo: Esta dimensión se le realizó una jerarquía con la estructura Modelo/Marca/Año.



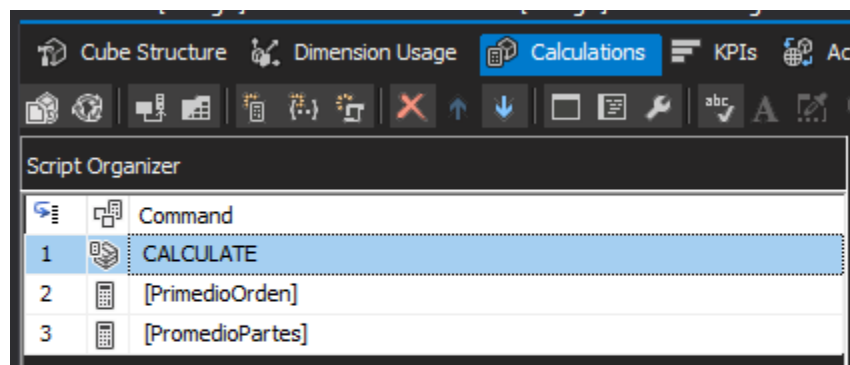
Dimensión Descuento: Esta dimensión se obtiene de la dimensión que se encuentra en el datawarehouse.



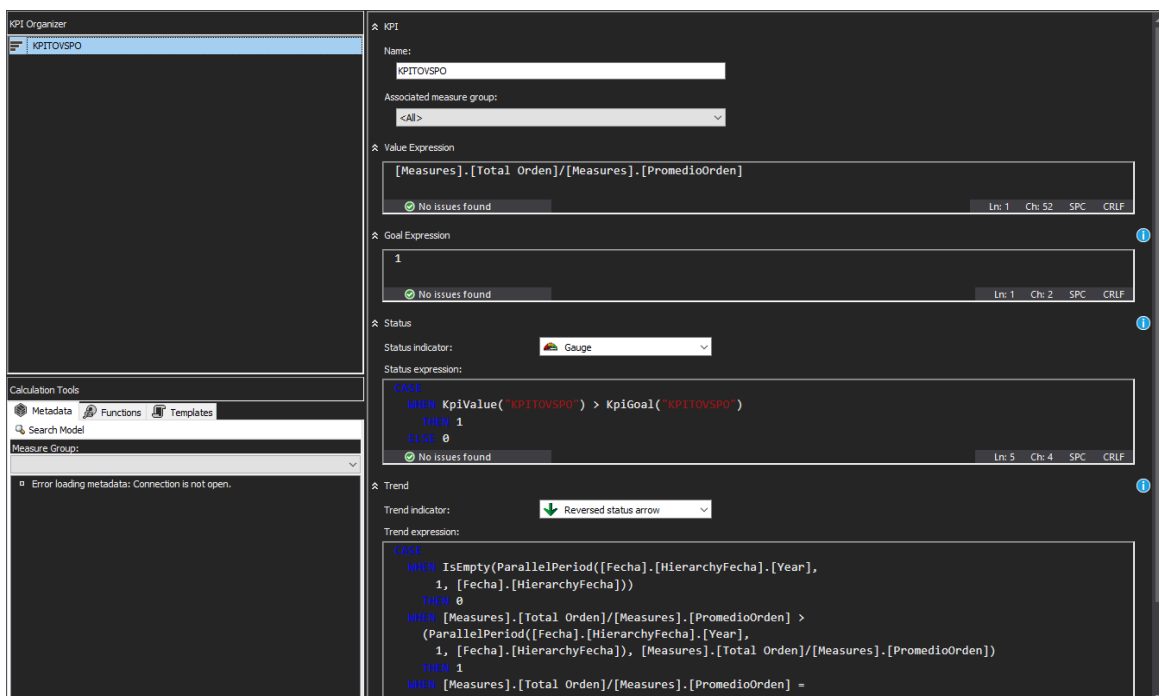
Dimensión Geografía: Esta dimensión se le realizó una jerarquía con la estructura Nombre País/Nombre Región/Nombre Ciudad.



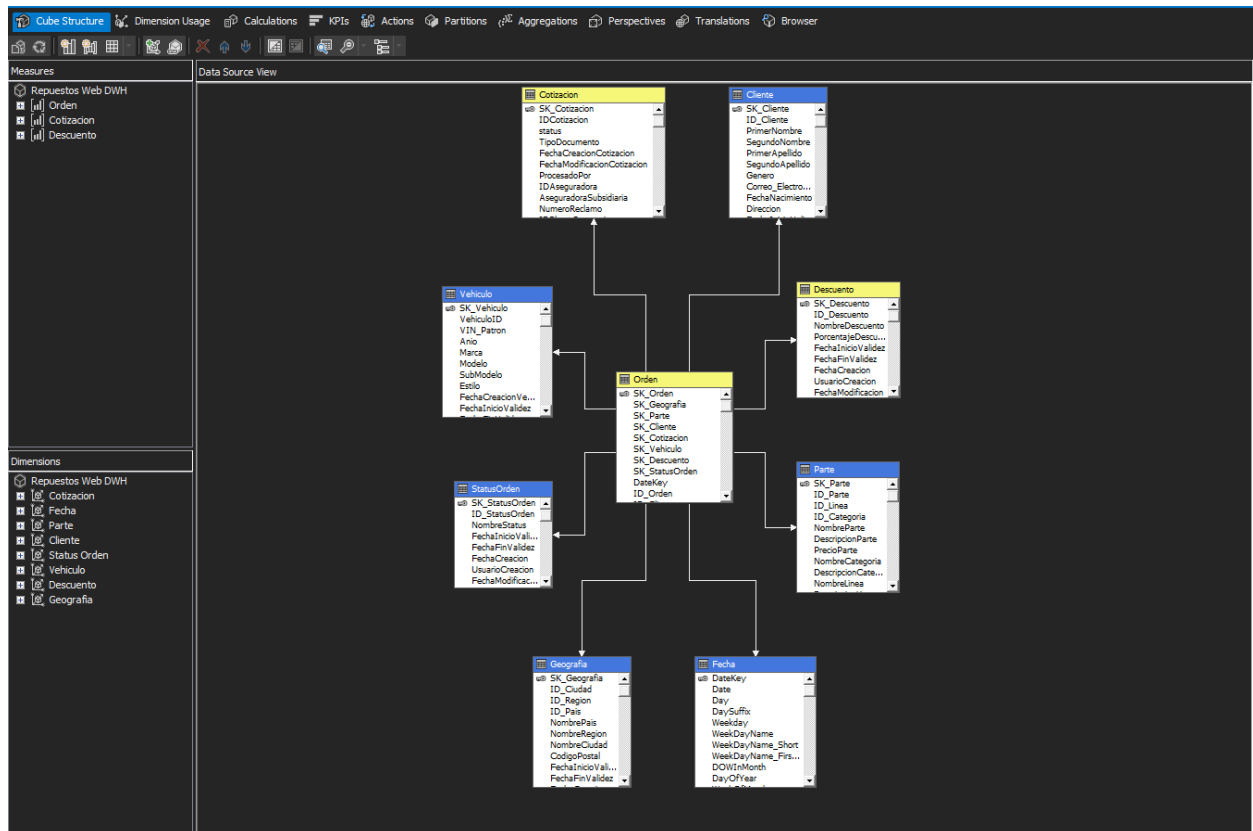
Columnas calculadas (Cubo): Columnas que originalmente no se tenían, pero se calcularon con la ayuda del cubo para poder utilizarlas en visualizaciones posteriores.



Columnas KPIs (cubo): En el cubo se crearon distintos KPIs para poder obtener un valor numérico, una especie de indicador que nos pueda decir si cierta cantidad se encuentra en un rango aceptable en base a nuestras necesidades.



Data source view: Aquí se puede observar todas las tablas que conforman nuestro cubo y como estas se relacionan entre sí.



JOB:

Después de tener los ETL se realizó un deploy a sql server donde se pudo programar para hacer que estos paquetes (ETLs) corrieran todos los días a las 11:55 PM.

