

FACULTAD DE INGENIERÍA
INGENIERÍA EN INFORMÁTICA Y SISTEMAS
ANÁLISIS DE DATOS I
CATEDRÁTICO STANLEY BOLAÑOS



Universidad
Rafael Landívar
Tradición Jesuita en Guatemala

MANUAL TÉCNICO

PROYECTO DEL CURSO

Lester Andrés García Aquino - 1003115
Mario Estuardo Gómez Orellana - 1020618
José Eduardo Tejeda León - 1097218
Mario Andrés Velasquez Banaay - 1092518

Guatemala, 25 de octubre de 2021

Índice

MODELO DE BASE DE DATOS DWH	3
Características	3
Descripción de las Tablas DWH	3
Aseguradoras	3
Clientes	3
Descuento	3
Fecha	3
Geografía	3
Partes	3
PlantaReparacion	4
StatusOrden	4
Vehiculo	4
OrdenCotizacion	4
Descripción de Esquemas DWH	4
Hecho	4
Dimensión	4
Staging	5
PROCESOS Y TAREAS ETL	6
Paquete de Integración SISS – Modelo Relacional	6
Flujo de Control del Proyecto	6
Aseguradora	7
Clientes	7
Descuento	8
Fechas	8
Geografía	9
Partes	10
Planta Reparación	11
Status Orden	11
Vehículo	12
REPORTERÍA EN POWER BI	20
OBJETOS VISUALES	20
JOBS DE SSMS	20

MODELO DE BASE DE DATOS DWH

Para la elaboración del esquema se tomó la decisión de realizar un modelo de estrella por la facilidad y comodidad de las tablas que la incorporan, bajo diferentes características:

Características

- El enfoque del trabajo está orientado a un mismo contexto que es almacenar los datos de Repuestos Web con énfasis en relaciones con aseguradoras las cuales solicitan cotizaciones para realizar órdenes de compra de forma centralizada.
- Todas las tablas de “*Dimensiones*” se relacionan a la tabla principal de “*Hecho.OrdenCotizacion*”, encargado de obtener toda la información relevante de las cotizaciones y ordenes realizadas por clientes registrados, aseguradoras o clientes sin registrar.
- El esquema de estrella es un esquema simple y fácil de entender, por lo tanto, la optimización de búsquedas es muy eficiente y facilita a su vez la generación de reportes, y al realizar consultas en base de datos, el ambiente se torna simple y sencillo al tomar como base la tabla de hechos.
- El cubo tiene la capacidad de crear jerarquías a partir de modelos estrella

Descripción de las Tablas DWH

Aseguradoras

Contiene los datos de las aseguradoras asociadas a las órdenes y cotizaciones de los repuestos web del sistema.

Clientes

Posee datos generales de los clientes que no son asociaciones ni empresas, únicamente personas.

Descuento

Contiene información relevante de los descuentos, usuarios creadores, porcentajes, etc.

Fecha

Posee diferentes niveles de especificidad respecto a tiempo, tales como día, mes, año, nombres de día, nombres de los meses, nombres de año, trimestres, y diferentes formatos para presentar rangos de tiempo.

Geografía

Almacena los datos específicos de dónde se realiza la orden, incluyendo ciudades, código postal, región, fechas de modificación y usuarios de creación.

Partes

Información enfocada en las partes a la venta, la línea y categoría que pertenece, precios y datos que se utilizan para búsquedas fáciles de partes.

PlantaReparacion

Contiene información de la ubicación de las plantas de reparación relacionadas con las partes, tales como la ciudad, país, estado de vigencia de activo o inactivo y nombre de la planta.

StatusOrden

Se encuentra con datos acerca del estado de la orden el cual puede actualizarse y mantener datos como última fecha de actualización.

Vehiculo

Identifica los vehículos registrados con referencia a datos como el patrón VIN, el año, modelo, estilo. Se relacionan directamente con las partes las cuales venden, siendo estas una referencia dentro de la tabla de partes.

OrdenCotizacion

Tabla principal de la base de datos, la cual contiene todas las referencias de las tablas mencionadas, teniendo referencias de la orden, cotización, vehículos, partes y datos específicos como el estado actual de la orden - cotización, direcciones de entrega, recotizaciones, rutas y cantidades.

Descripción de Esquemas DWH

Hecho

Hace referencia a las tablas que poseen valores de referencia y específicos de las órdenes y de las cotizaciones, el cual únicamente contiene una tabla, definida con el nombre de “*OrdenCotizacion*” en la cual acepta relaciones entre clientes, aseguradoras y clientes no registrados con órdenes de compra y cotizaciones realizadas.

Dimensión

Referente de las tablas que contienen información especificada por secciones que desnormaliza tablas para poder eficientar el proceso de cargas a cubos. Las tablas que se encuentran relacionadas a este esquema son:

- Dimensión Aseguradoras
- Dimensión Clientes
- Dimensión Descuento
- Dimensión Fechas
- Dimensión Geografía
- Dimensión Partes
- Dimensión Planta de Reparación
- Dimensión Status de Orden
- Dimensión Vehículo

Staging

Hace referencia a las tablas que poseen un valor cualitativo y cuantitativo dentro del modelo data warehouse, con el fin de contar con tablas temporales para poder realizar cálculos previos y analizar estadísticamente la información obtenida. Las tablas que se encuentran relacionadas a este esquema son:

- Staging Orden de Cotización

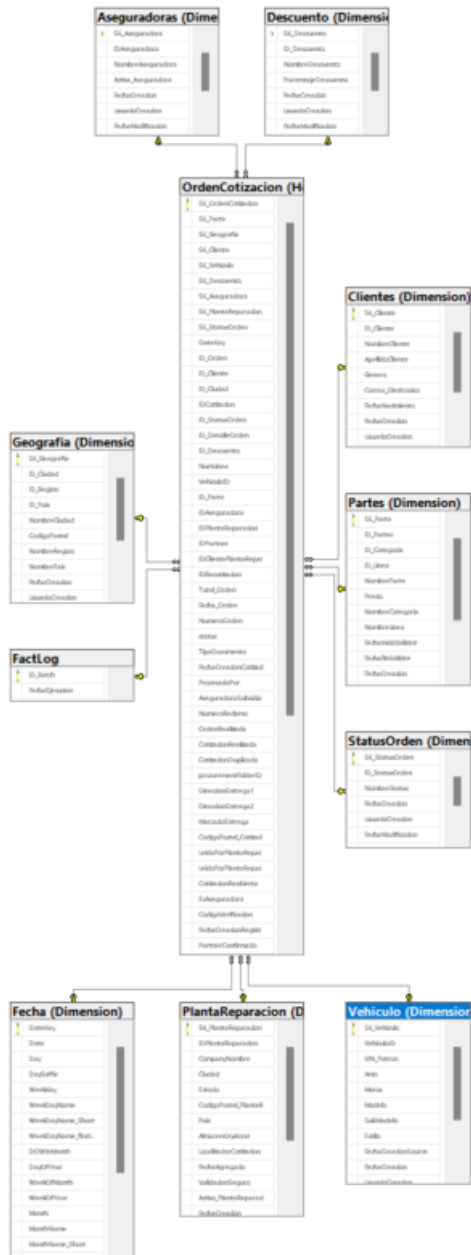


Imagen 1: Modelo Data Warehouse – Modelo de Estrella

PROCESOS Y TAREAS ETL

Paquete de Integración SISS – Modelo Relacional

Para realizar la carga de información se toma la decisión de utilizar un proceso ETL debido a las facilidades de la herramienta para cumplir con los distintos requisitos de la elaboración del proyecto, para poder transformar los datos y manipular dicha información antes de pasar al proceso Data Warehouse, ya que se utilizó un modelo ROLAP lo que requirió que los datos fuesen procesados antes de llegar a la implementación lógica del cubo y Data Warehouse. Además, al implementar un proceso de carga en archivos .CSV, siendo fuentes distintas de datos, se prosiguió a transformar los datos con el fin de estandarizar estos en caso existan ciertas inconsistencias que pudieran afectar la carga de datos.

El modelo ROLAP fue utilizado por motivos de eficiencia respecto a tiempos ya que no se contó con el suficiente poder computacional para procesar la cantidad de datos requeridos en tiempo. Esto también fue con el fin de evitar un gasto de recursos innecesarios.

El paquete ETL se encarga de desnormalizar y consolidar datos que se encuentran almacenados en archivos SQL para poder transformarlos y utilizar únicamente los datos necesarios. Por medio del proceso ETL (Extract – Transform – Load) se ejecutó este proceso.

Flujo de Control del Proyecto

Para la solución del proyecto se procede iniciando con la creación de la base de datos Data Warehouse correspondiente por medio de un script diseñado para implementar el modelo estrella y posteriormente transformar los datos actuales. En dicha transformación se ejecuta un proceso de Slowly Changing Dimension para cada dimensión con el fin de unir datos de la fuente de la base de datos OLTP y el archivo .CSV al destino final en Data Warehouse.

Previo al proceso de Slowly Changing Dimension, se ejecuta la obtención de los datos de las fuentes y se realiza un proceso de unión para posteriormente agregar columnas derivadas las cuales son las columnas de linaje y auditoría. Esto teniendo los siguientes flujos:

Aseguradora

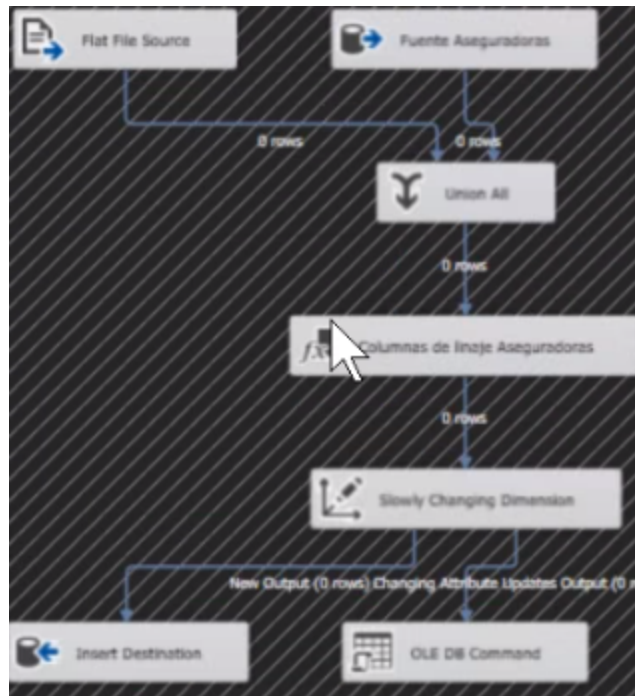


Imagen 2: Flujo de control de Aseguradora

Clientes

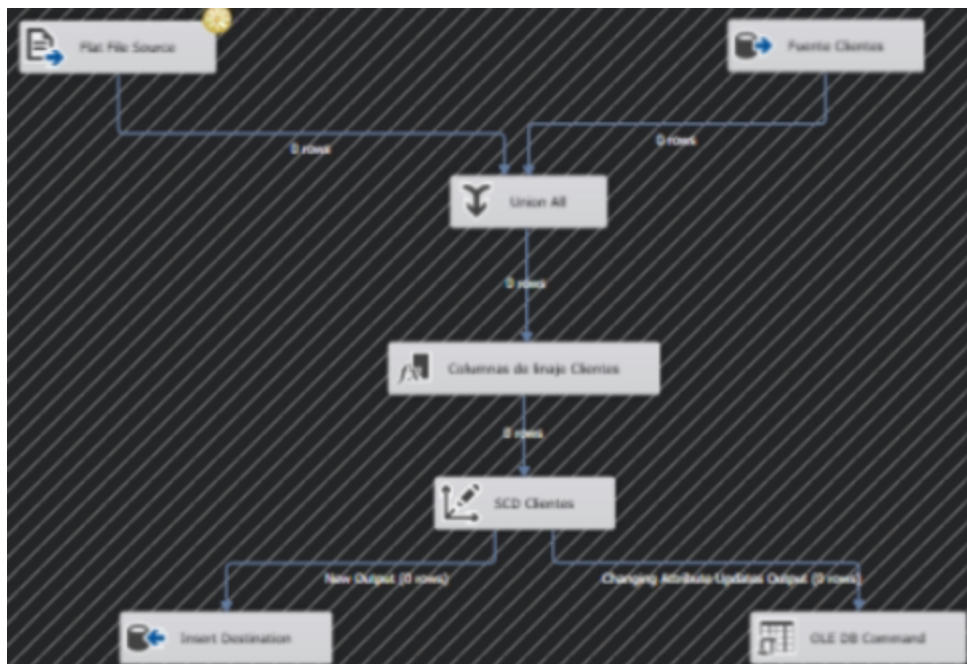


Imagen 3: Flujo de control de Clientes

Descuento



Imagen 4: Flujo de control de Descuento

Fechas

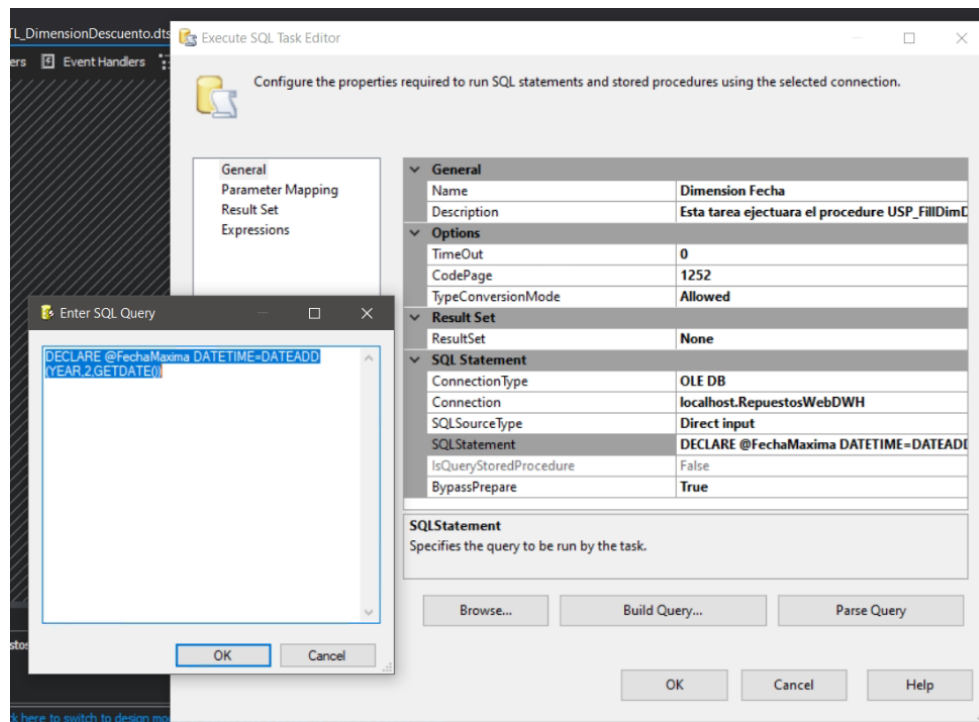


Imagen 5: Flujo de control de Fechas

Geografía

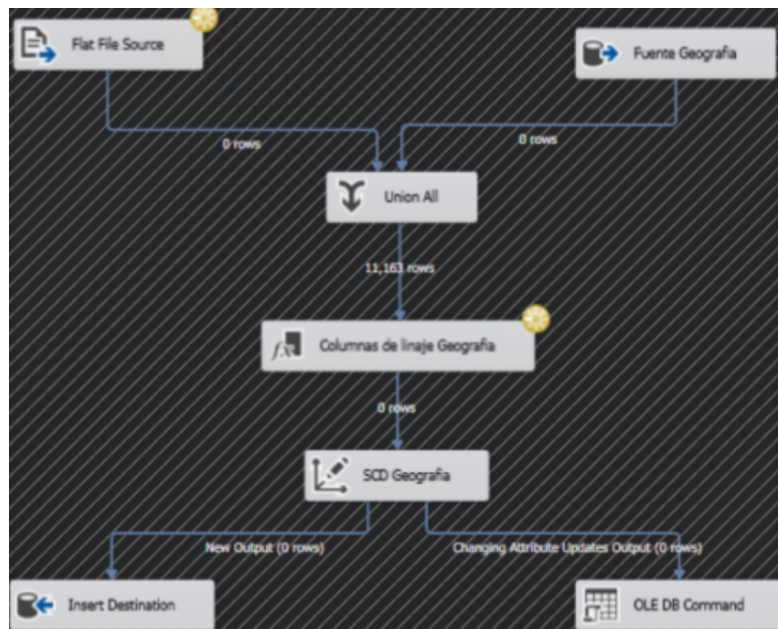


Imagen 6: Flujo de control de Geografía

Partes

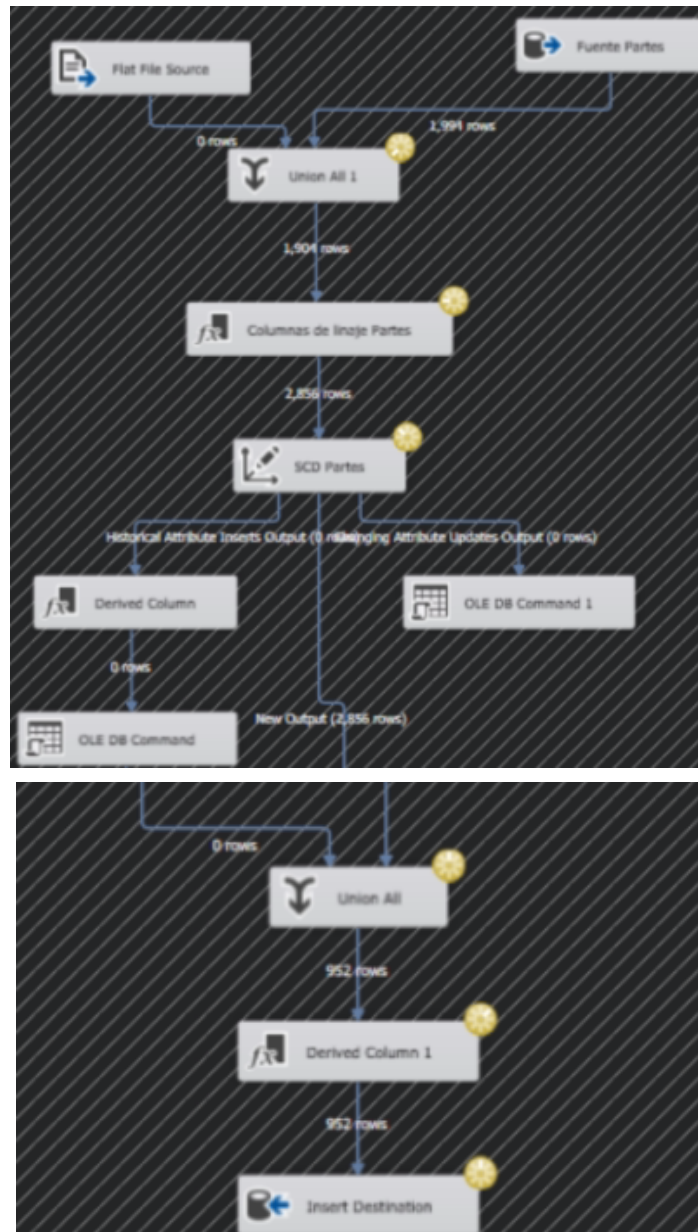


Imagen 7: Flujo de control de Partes

Planta Reparación

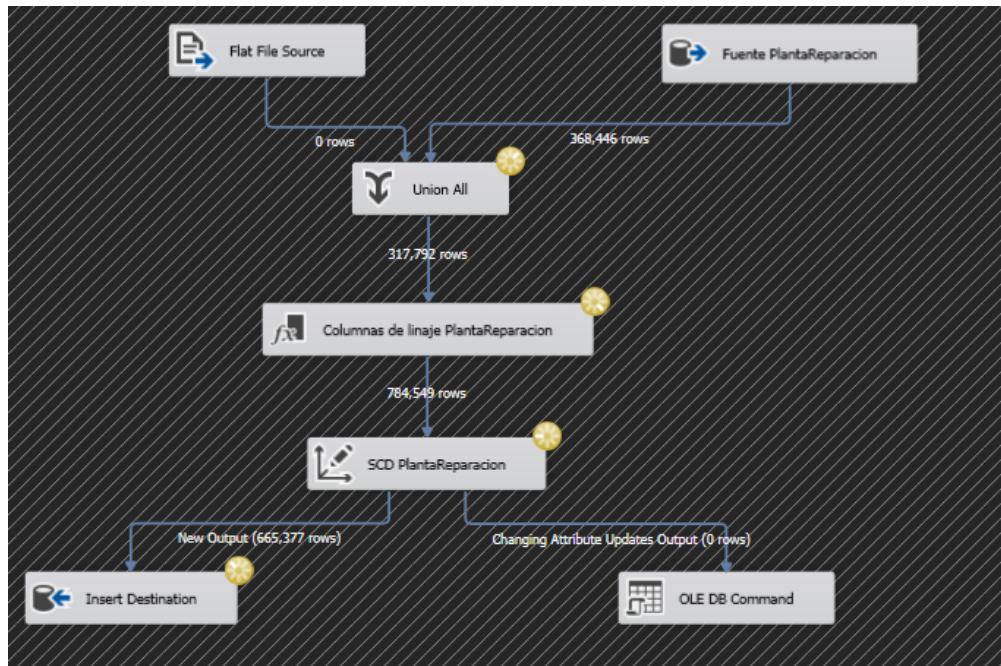


Imagen 8: Flujo de control de Planta de Reparación

Status Orden

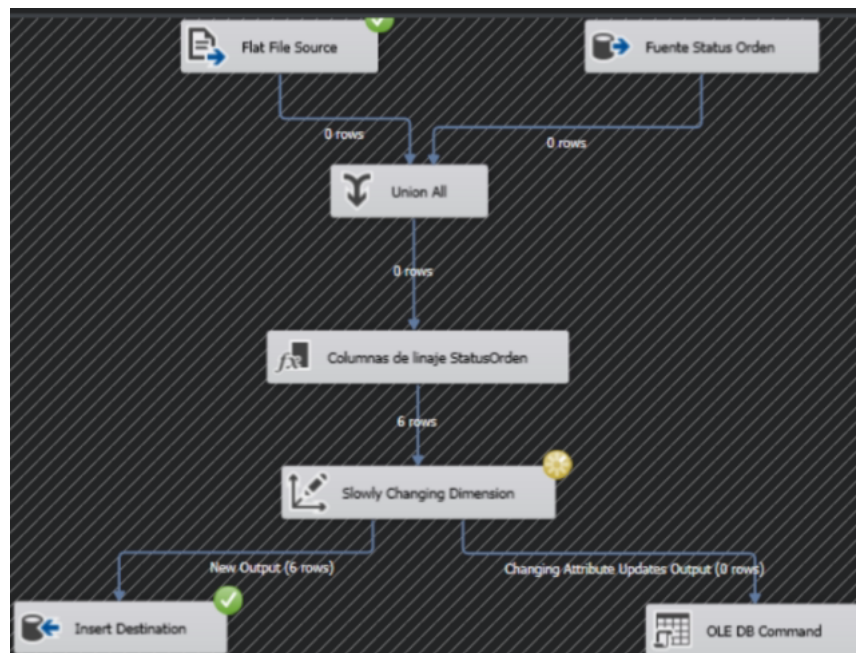


Imagen 9: Flujo de control de Status Orden

Vehículo

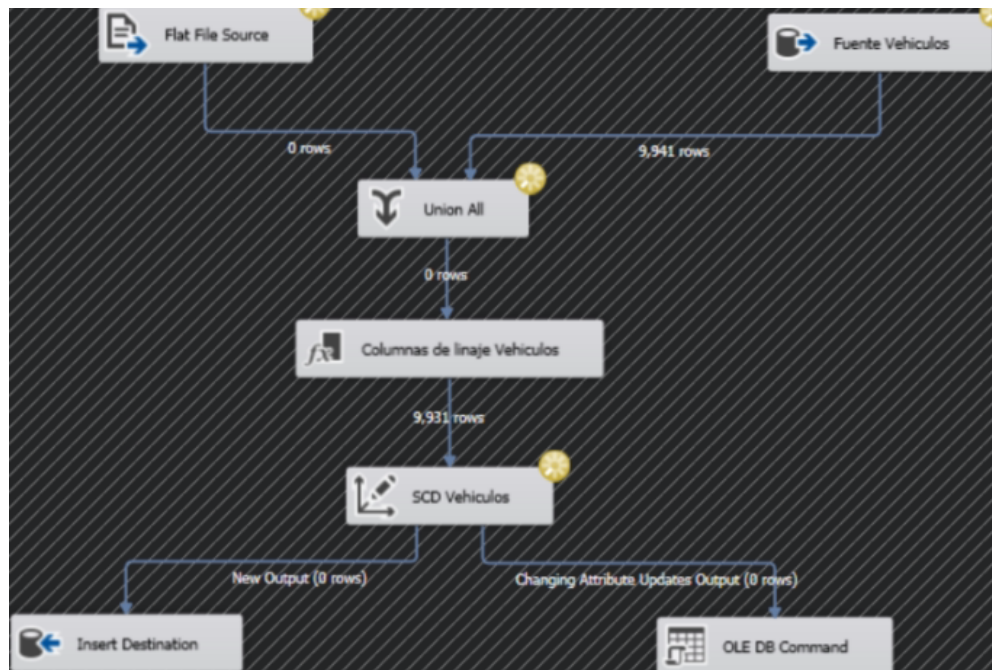


Imagen 10: Flujo de control de Vehículos

Siguiente a esto se procesa la tabla de hechos a base de lo procesado anteriormente en las dimensiones, lo cual lleva al siguiente flujo de operaciones:

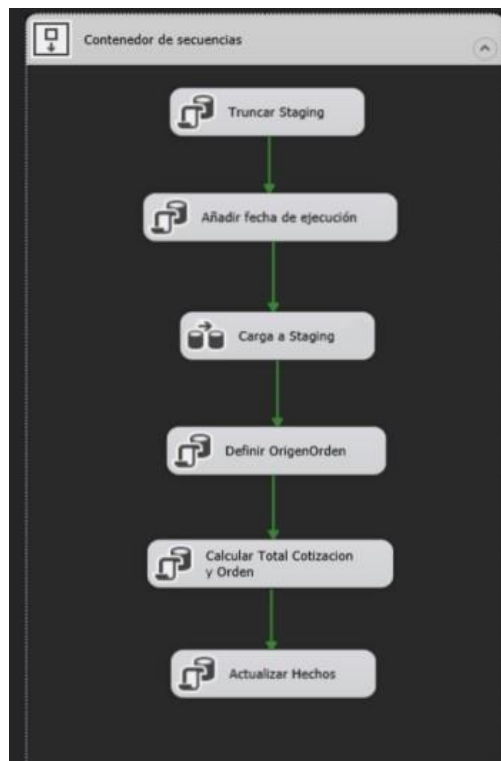


Imagen 11: Flujo de control de Hechos

El proceso incluye un traslado de datos al área de staging consiste en truncar dicha tabla inicialmente, con el fin de tener un traslado de datos limpio. Siguiendo a esto se ejecuta un SQL Task que ejecuta un Stored Procedure el cual asigna al FactLog los datos necesarios como el ID_Batch y ID_SourceSystem para llevar registro de qué datos utilizar y almacenar datos del procesamiento actual.

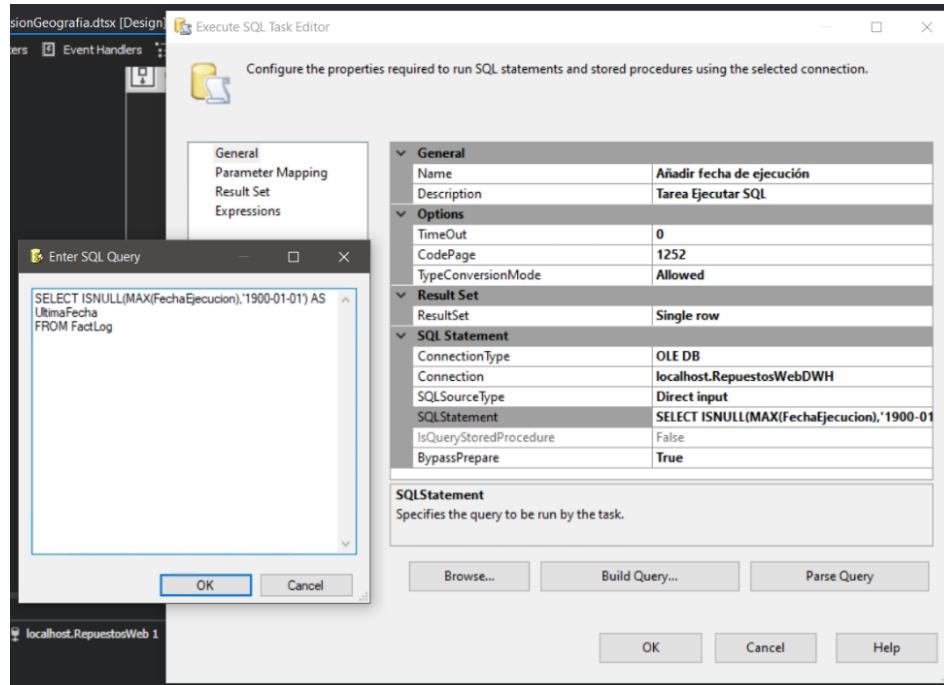


Imagen 12: Query ejecutado para procesar fechas nuevas

Para la carga a Staging se sigue el siguiente flujo de unión hacia dicha tabla:

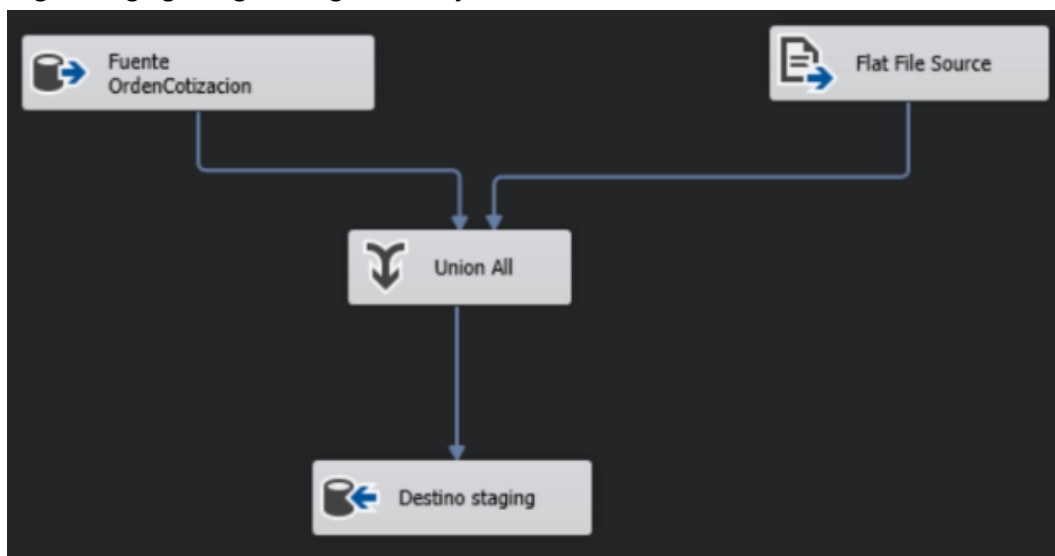


Imagen 13: Unión de fuentes de datos para staging

Luego del traslado a Staging, se procede a definir la columna OrigenOrden, el cual es por medio de la ejecución de un UPDATE realizando el cálculo mostrado en la siguiente consulta:

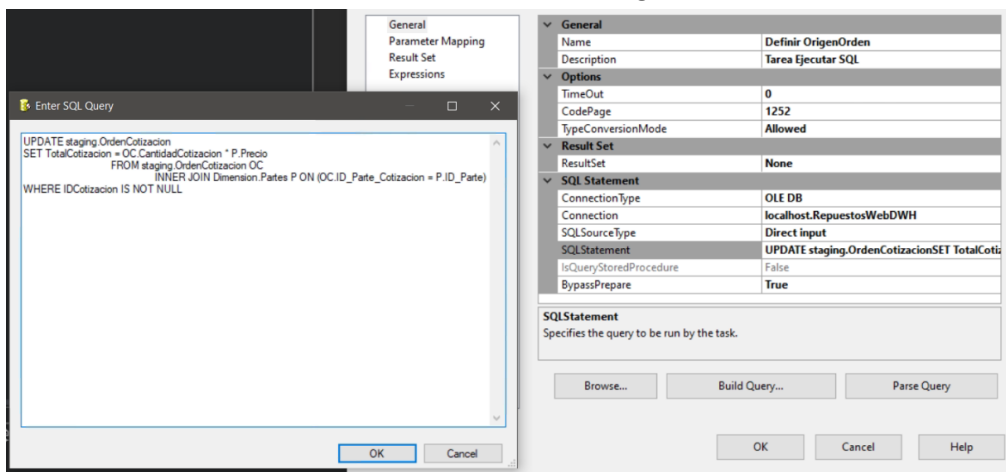


Imagen 14: Query para actualizar los datos de origen

Llegando a los últimos pasos, se procede a calcular el total de cotizaciones por medio del siguiente proceso el cual almacena los datos en la columna TotalCotizacion, para su futuro uso dentro del cubo a implementar:

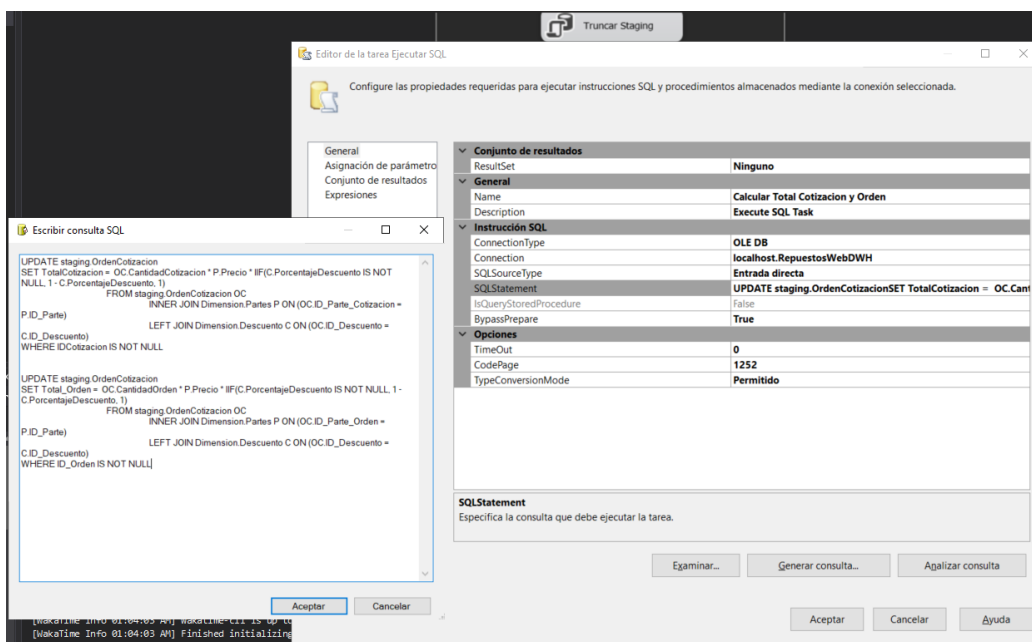


Imagen 15: Cálculo de totales de cotización y orden

Finalmente, se trasladan los datos ya transformados y procesados a la tabla final de Hechos la cual es por medio de la ejecución de un Stored Procedure que realiza un Merge entre la tabla de Staging.OrdenCotizacion y Hecho.OrdenCotizacion. Este proceso se visualiza de la siguiente forma:

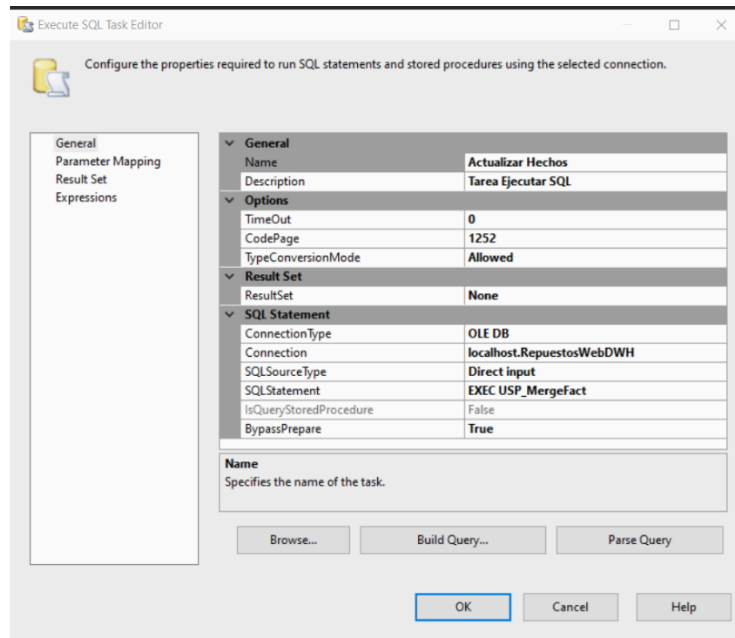


Imagen 16: Ejecución de Stored Procedure para merge de datos a tabla de Hechos

Para centralizar la ejecución de todos los procesos en un mismo lugar, se hizo uso de un nuevo paquete el cual agrupa la ejecución de todas las dimensiones a transformar, para luego transformar los datos a la tabla de hechos.



Imagen 17: Centralizado de ejecución de paquetes

DESARROLLO DE CUBO

El desarrollo del cubo consistió en realizar la conexión a la base de datos modelo Data Warehouse para exportar dimensiones y hechos con el fin de poder acceder a este y realizar mediciones las cuales serán procesadas y desplegadas más adelante por medio de la herramienta de Power BI.

Inicialmente, la conexión a Data Warehouse permitió la visualización gráfica de las tablas que serían utilizadas, desplegando lo siguiente:

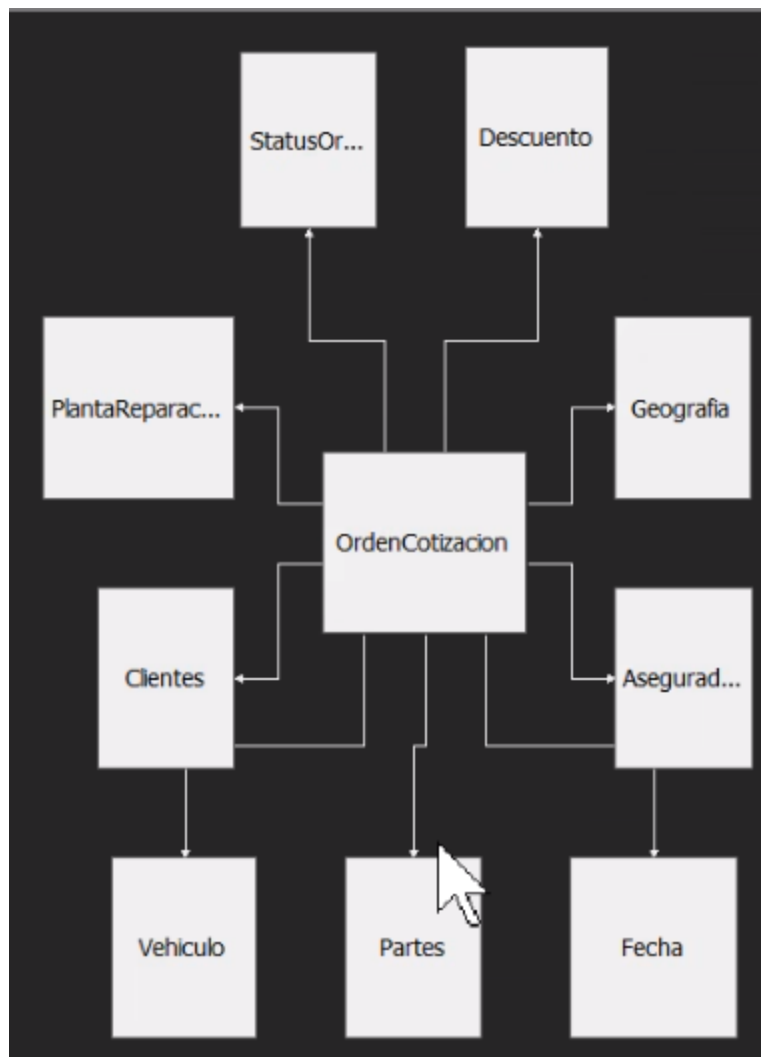


Imagen 18: Modelo Data Warehouse en SSAS

Al momento de tener acceso al modelo de Data Warehouse, se generan las jerarquías necesarias en caso existan para poder realizar *roll up* y *drill down* durante el acceso a los datos más adelante.

Para realización de jerarquías se enlistan las columnas de los datos a desplegar dentro de un proyecto SSAS generado en Visual Studio, lo cual permitió de forma más clara y fácil de implementar, un modelo

jerárquico definido en la ventana ubicada en el medio de la siguiente imagen:

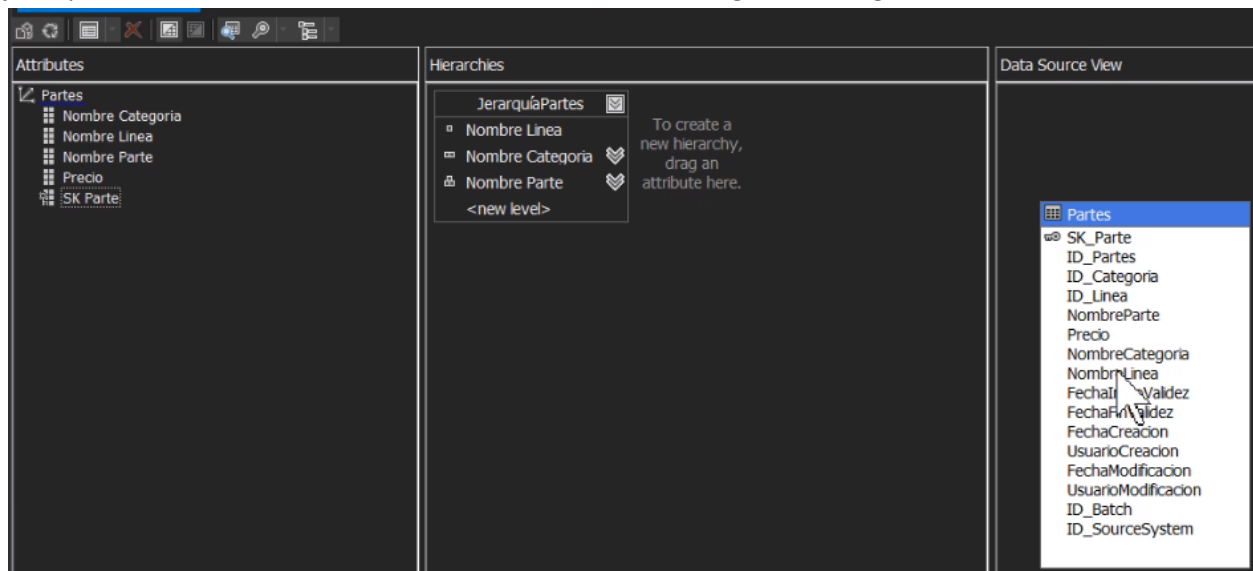


Imagen 19: Configuración de jerarquías para tabla Partes

Siguiente a este proceso, se realiza la relación de atributos a la llave subrogada con los atributos conectados de menor jerarquía a mayor jerarquía de forma continua. En caso que alguna columna no pertenezca a alguna jerarquía, se asocia directamente a la llave subrogada.

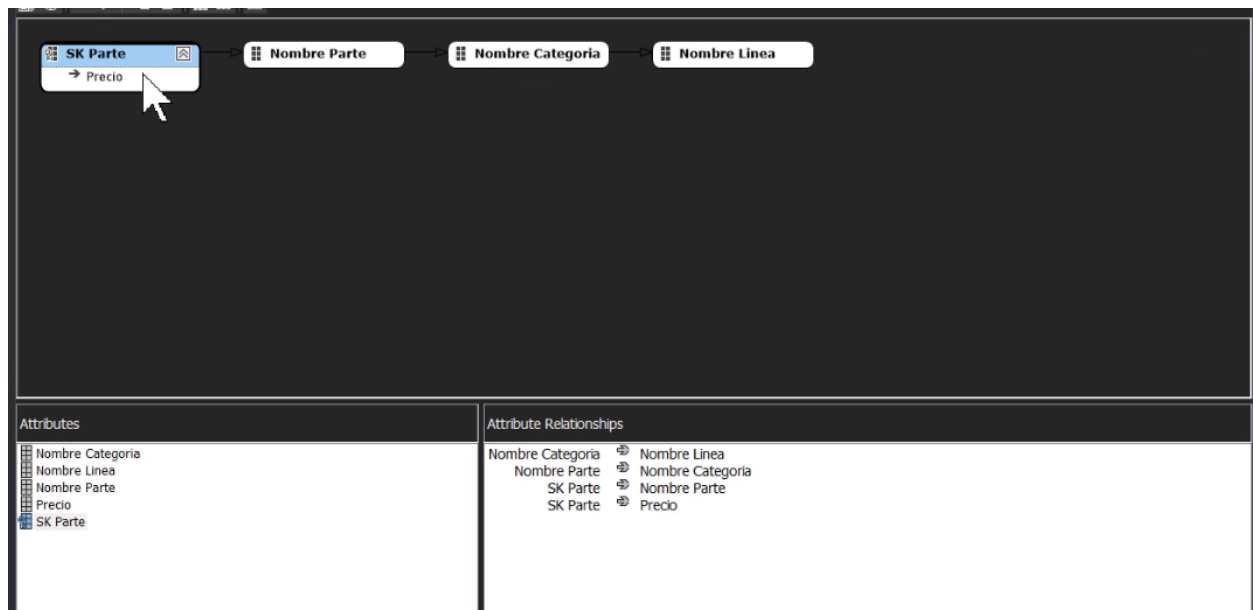


Imagen 20: Asociación y relación entre atributos con y sin jerarquía

Teniendo implementadas las jerarquías necesarias, se procede a realizar las medidas básicas del cubo, las cuales son generadas por medio de funciones de agregación. Adelante tenemos un ejemplo de algunas medidas generadas.

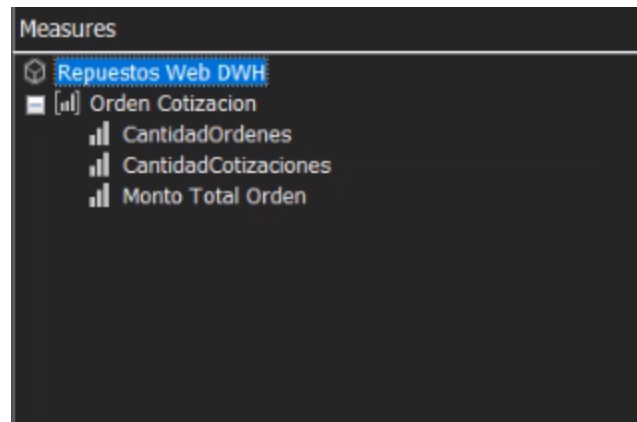


Imagen 21: Medidas del proyecto

Por medio de las medidas creadas a base de funciones de agregación, se generan las medidas personalizadas necesarias con cálculos más específicos indicadas en el espacio de “Expression”.

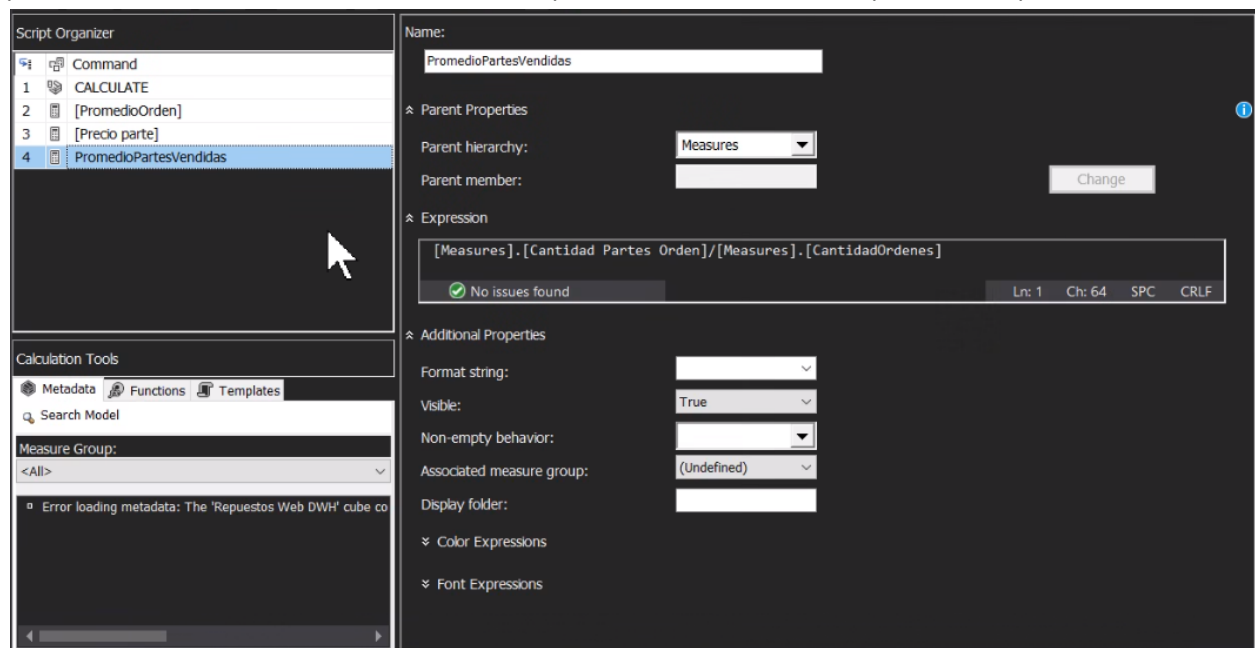


Imagen 22: Configuración de medidas personalizadas

Para finalizar los últimos datos a desplegar en el cubo, se generan los indicadores clave (KPI) que indican el estado de cada dato a medir relacionado con este indicador. Estos utilizan un Status expression como la siguiente imagen, la cual es utilizada para desplegar de forma más visual el estado actual del dato medido y comparado con el KPI.

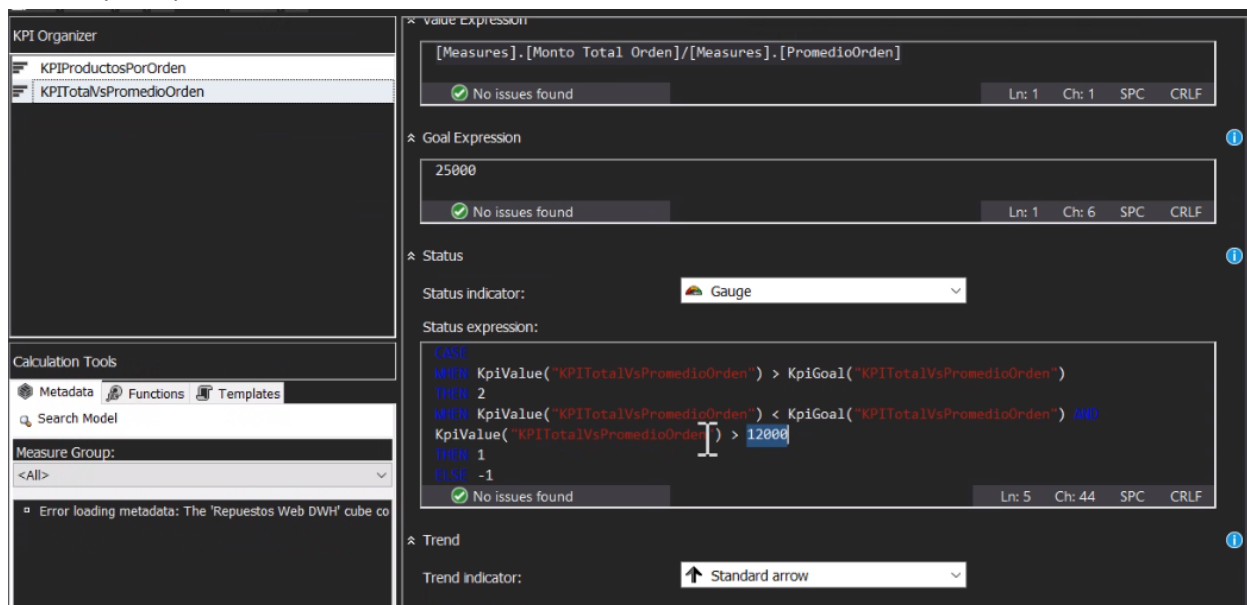


Imagen 23: Configuración general de KPI KPITotalVsPromedioOrden

Además, se implementó la función de tendencias, las cuales utilizan el KPI para poder obtener predicciones y tendencias de parte de los datos históricos de este indicador. Para configurar esta interacción se utiliza el siguiente flujo:

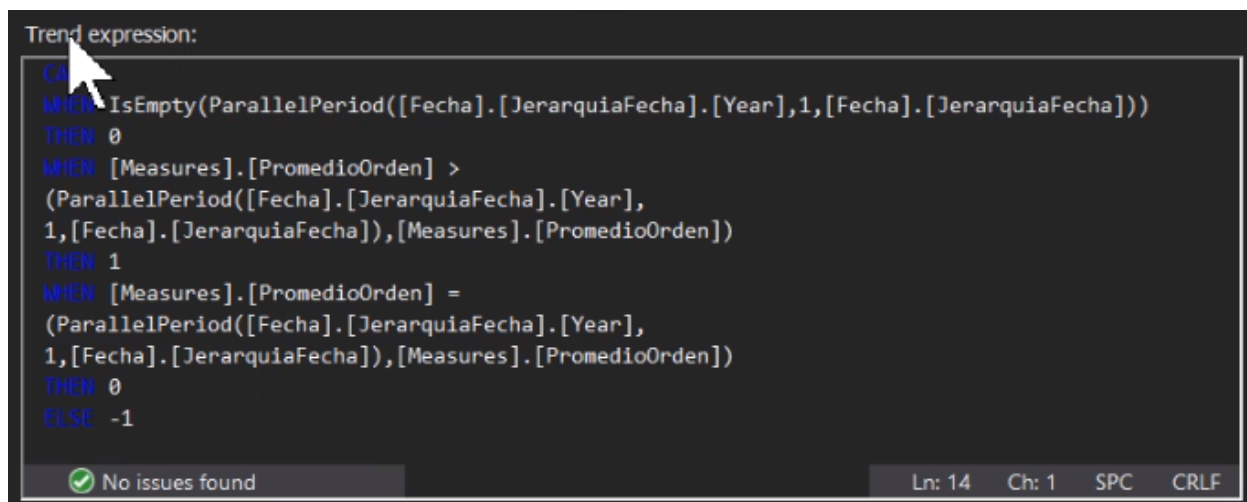


Imagen 24: Configuración de definición de tendencias para Partes

Para permitir un proceso fluido y sin errores, basado en los puntos necesarios del proyecto y la eficiencia respecto al tiempo, se utilizó un modo de almacenamiento ROLAP.

Elem...	Nombre de partición	Origen	Filas estimadas	Modo de almacenamiento	Diseño de agregaciones
1	Orden Cotización	OrdenCotizacion	0	ROLAP	




Imagen 25: Configuración de Modo de Almacenamiento

REPORTERÍA EN POWER BI

Para el despliegue de reportería por medio de la herramienta Power BI, se hizo uso de la conexión al servicio de SSAS junto con la instalación de objetos visuales adicionales para una mejor interpretación y visualización más amigable de los datos obtenidos.

OBJETOS VISUALES

Para la conexión al cubo implementado en el servicio SSAS, se procedió al acceso de la tienda de Power BI con el fin de instalar los siguientes objetos visuales extras:

- ArcGIS Maps for Power BI 
- Forecasting TBATS 
- Power KPI Matrix 3.1.1 

JOBS DE SSMS

Se hizo una conexión al servicio SSAS, el cual consistió previamente en la creación de Jobs desde *Integration Services Catalogs* en el cual se hizo la conexión al proyecto de SSIS para poder automatizar la ejecución de paquetes por medio de SQL Server Agent. El Job fue nombrado *ActualizaRepuestosWeb* que contiene la ejecución de transformación de dimensiones y hechos para proceder a actualizar el cubo implementado.



Imagen 26: Job encargado de ejecutarse diariamente

Propiedades del trabajo - JobProyectoETLRepuestosWebDWH				
<div> <div> <div>Seleccionar una página</div> <div> <div>General</div> <div>Pasos</div> <div>Programaciones</div> <div>Alertas</div> <div>Notificaciones</div> <div>Destinos</div> </div> </div> <div> <div>Script</div> <div>Ayuda</div> </div> </div>				
Lista de pasos de trabajo:				
Paso	Nombre	Tipo	En caso de éxito	En caso de error
1	Actualizar dimensiones y hechos	Paquete SQL Server Integration Services	Salir del trabajo e informar del éxito	Salir del trabajo e informar del error

Imagen 27: Pasos de ejecución del Job

Por motivos de automatización como se mencionó anteriormente, se procedió a implementar la ejecución todos los días a las 10PM, siendo enlistada en el apartado de Schedules.

Propiedades del trabajo - JobProyectoETLRepuestosWebDWH

Seleccionar una página	Script	Ayuda
General		
Pasos		
Programaciones		
Alertas		
Notificaciones		
Destinos		

Lista de programaciones:

ID	Nombre	Habilitado	Descripción	Trabajos en programa...
Nuevo	Diariamente a las 12 am	Si	Sucede cada día a las 12:00:00. Se utilizará la programación que empieza el 25/10/2021.	

Imagen 28: Ejecución automatizada diariamente a las 10PM