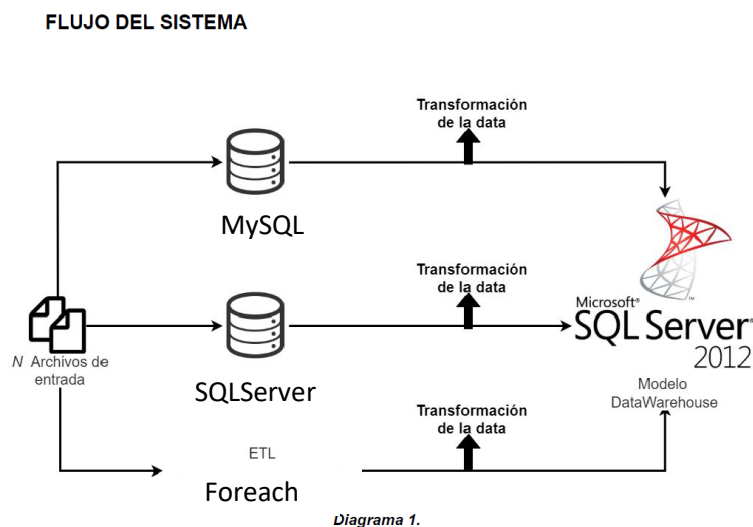




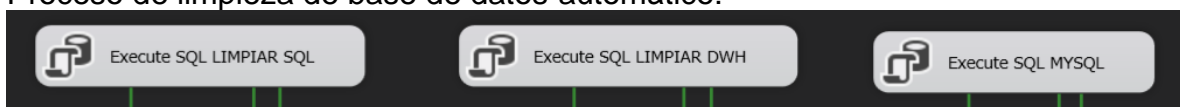
Proyecto 1 Fase 1

“GuateFood”

Descripción de los pasos que se tomaron desde la extracción de los datos hasta la carga al Datawarehouse:



Proceso de limpieza de base de datos automático:



Execute SQL es utilizado para ejecutar Scripts directos hacia cada base de datos utilizada, en el proceso de ETL.

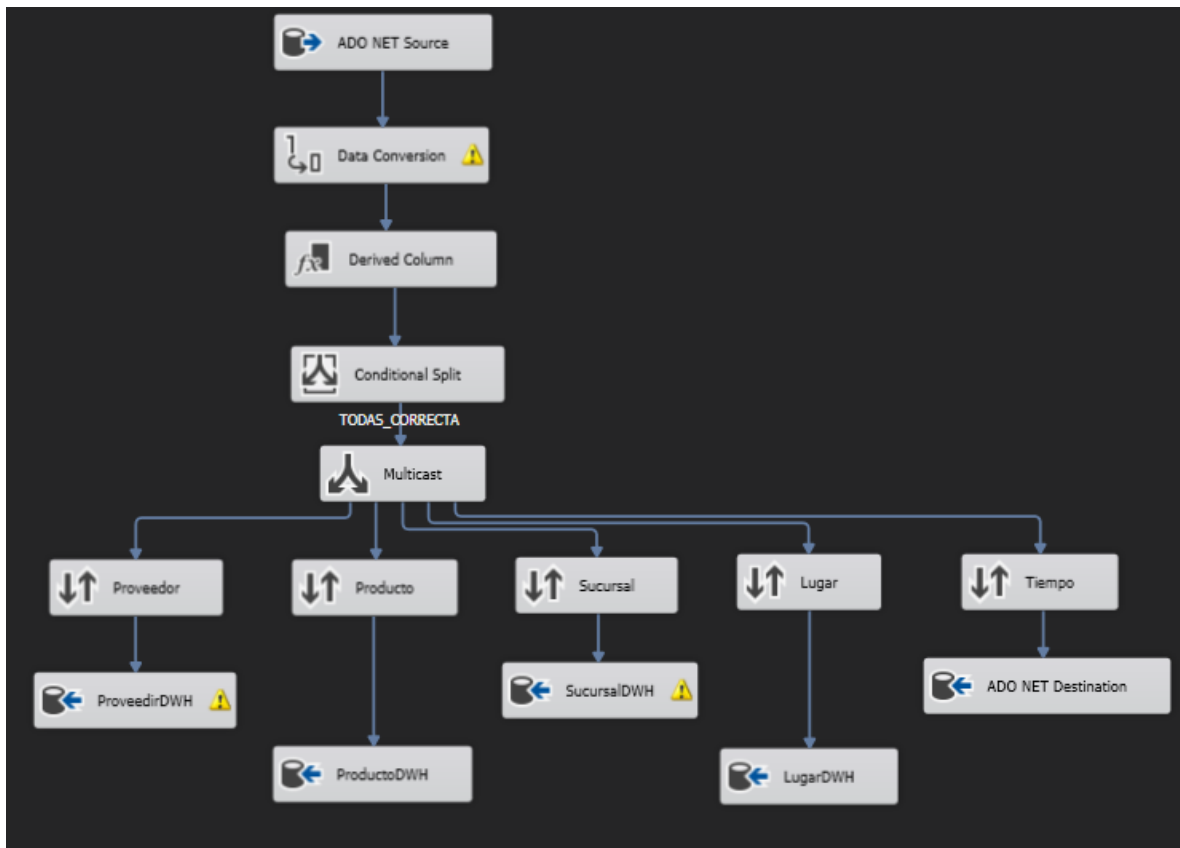
Proceso utilizado donde se implementa la base de datos MySQL y SQL Server



Llenar Temporal MYSQL (TaskScript):

Donde se escribe el código para leer el archivo e ir a insertar a la tabla temporal de compras y ventas que usa como pivote para hacer el proceso de ETL y luego llenar el DWH.

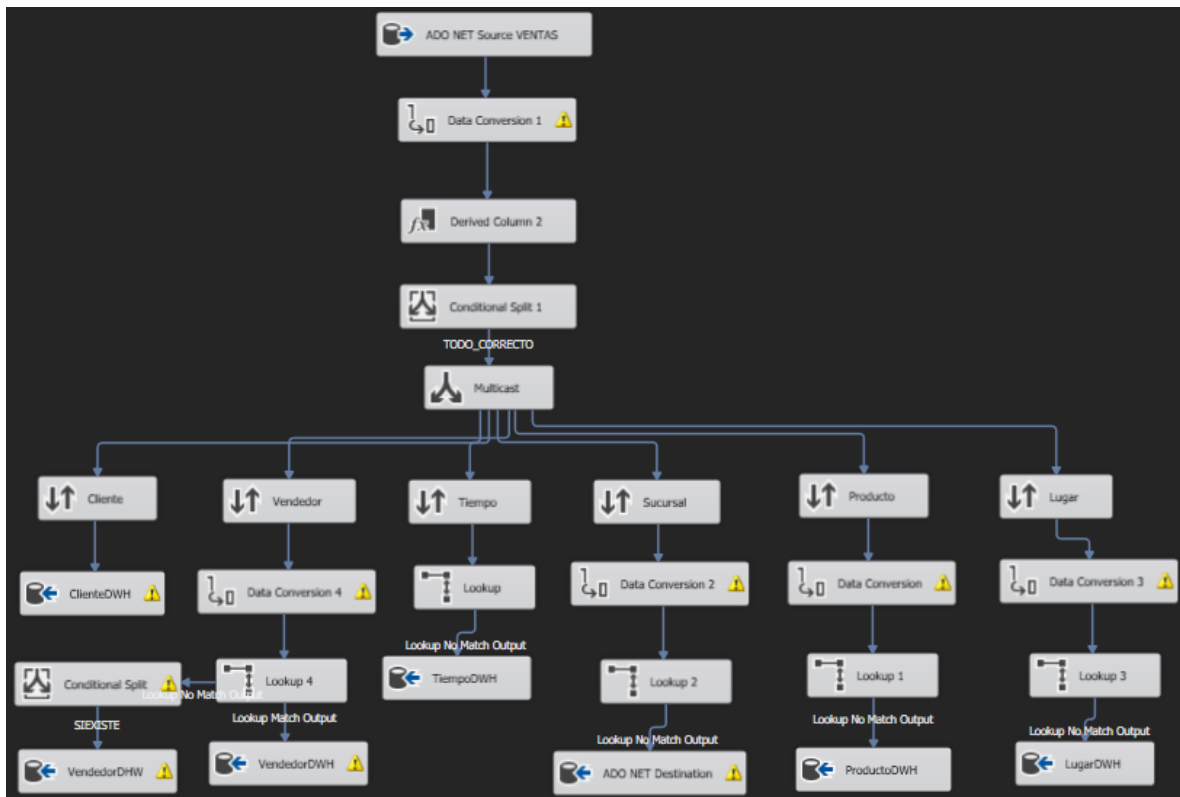
Llenar de Compras 1(Data Flow Task):



Hace el proceso de limpieza de datos y conversión de datos al tipo que se va a insertar su respectiva tabla del DWH.

- ADO NET Source: vamos a la tabla temporal de compras a extraer los datos.
- Data Conversion: Transforma el tipo de datos y limpia los que no cumplen para la transformación.
- Derived Colum: Utilizado para validar si los datos son nulos o no son nulos
- Conditional Split: se limpia la data respecto a la condición si es nula o no y se dejan como salida solo todas las filas correctas.
- Multicast: es para redirigir el flujo de datos y que en todos los sort haya la misma cantidad de datos de entrada.
- Sort(Proveedor, Producto, Sucursal, Lugar, Tiempo): utilizado para quitar los datos duplicados respecto al codigo de cada una de las tablas.
- ADO Net Destination / OLE DB Destination: utilizado para conectar a la base de datos y así hacer las inserciones.

Llenar de Ventas 1:



Hace el proceso de limpieza de datos y conversión de datos al tipo que se va a insertar su respectiva tabla del DWH.

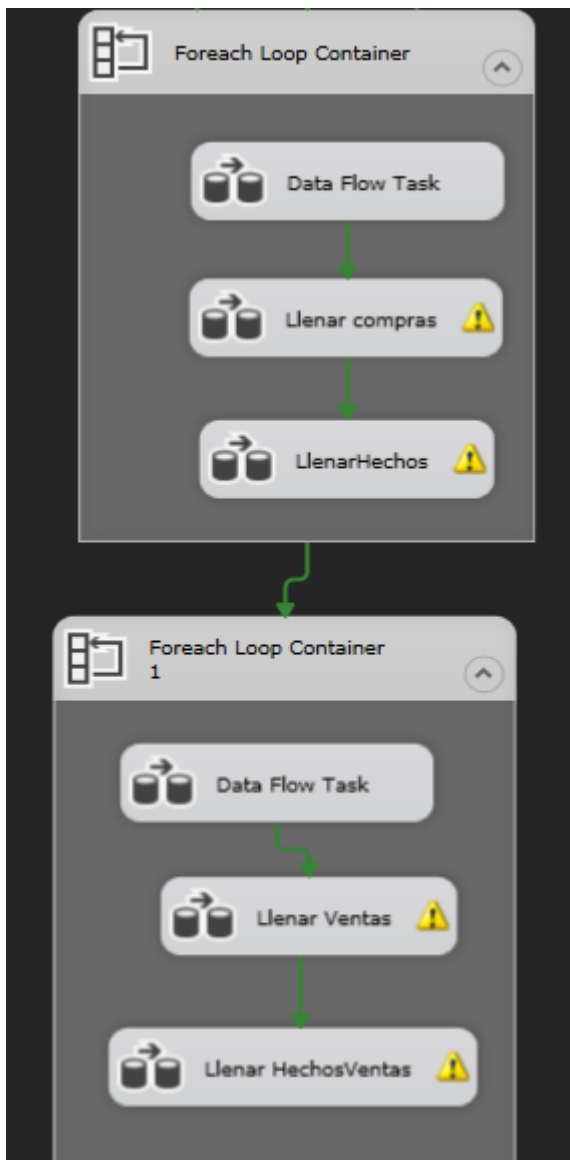
- ADO NET Source: vamos a la tabla temporal de ventas a extraer los datos.
- Data Conversion: Transforma el tipo de datos y limpia los que no cumplen para la transformación.
- Derived Colum: Utilizado para validar si los datos son nulos o no son nulos
- Conditional Split: se limpia la data respecto a la condición si es nula o no y se dejan como salida solo todas las filas correctas.
- Multicast: es para redirigir el flujo de datos y que en todos los sort haya la misma cantidad de datos de entrada.
- Sort(Cliente, Vendedor, Tiempo, Sucursal, Producto, Lugar): utilizado para quitar los datos duplicados respecto al codigo de cada una de las tablas.
- ADO Net Destination / OLE DB Destination: utilizado para conectar a la base de datos y así hacer las inserciones.
- LookUp: es una herramienta para buscar coincidencias en las tablas y así no insertar data repetida.

Llenar Hechos:



- ADO NET Source (Compras/Ventas): es un acceso a las tablas temporales y así poder comparar el registro que hay que de la respectiva venta o compra .
- Data Conversión: cambia el tipo de dato.
- Derived Colum: Utilizado para validar si los datos son nulos o no son nulos
- Conditional Split: se limpia la data respecto a la condición si es nula o no y se dejan como salida solo todas las filas correctas.
- LookUp: se hace el lookup para buscar las condincidencia de los registro que se tienen en las tablas del DWH y así devolver el id de cada tabla para usarla como llave subrogada en la tabla de hechos ventas y tabla de hecho compras y por último se inserta en su respectiva tabla con el elemento ADO Net Destination / OLE DB Destination.

Proceso utilizado donde se implementa el ForEach:

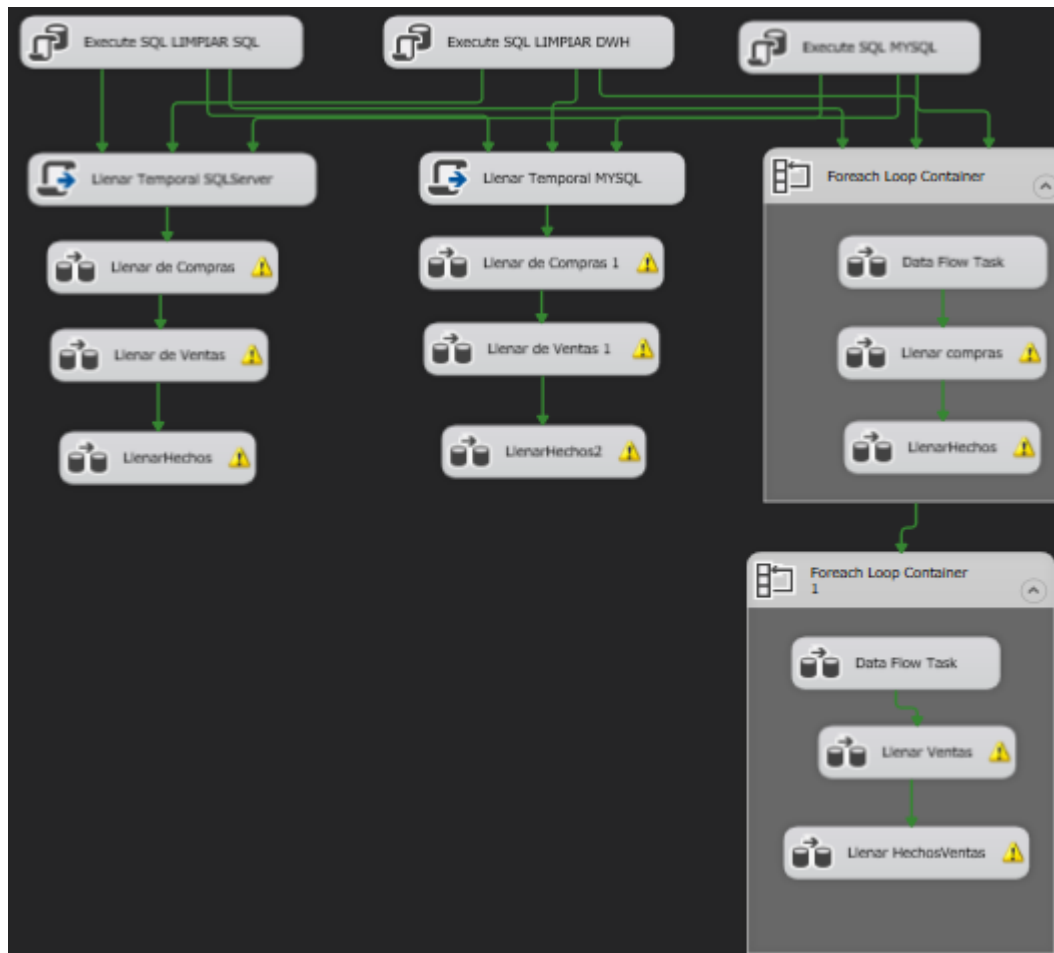


- Foreach Loop Container: utilizado para la lectura de archivos.
- Data Flow Task: se utiliza para cargar la data del archivo leído en la tabla temporal compras.
- Llenar Compras: es el proceso de llenado de tablas que van a formar luego la tabla hechos compras.
- Llenar Hechos: es el proceso que se utiliza para llenar la tabla de hecho compras con el proceso descrito anteriormente.

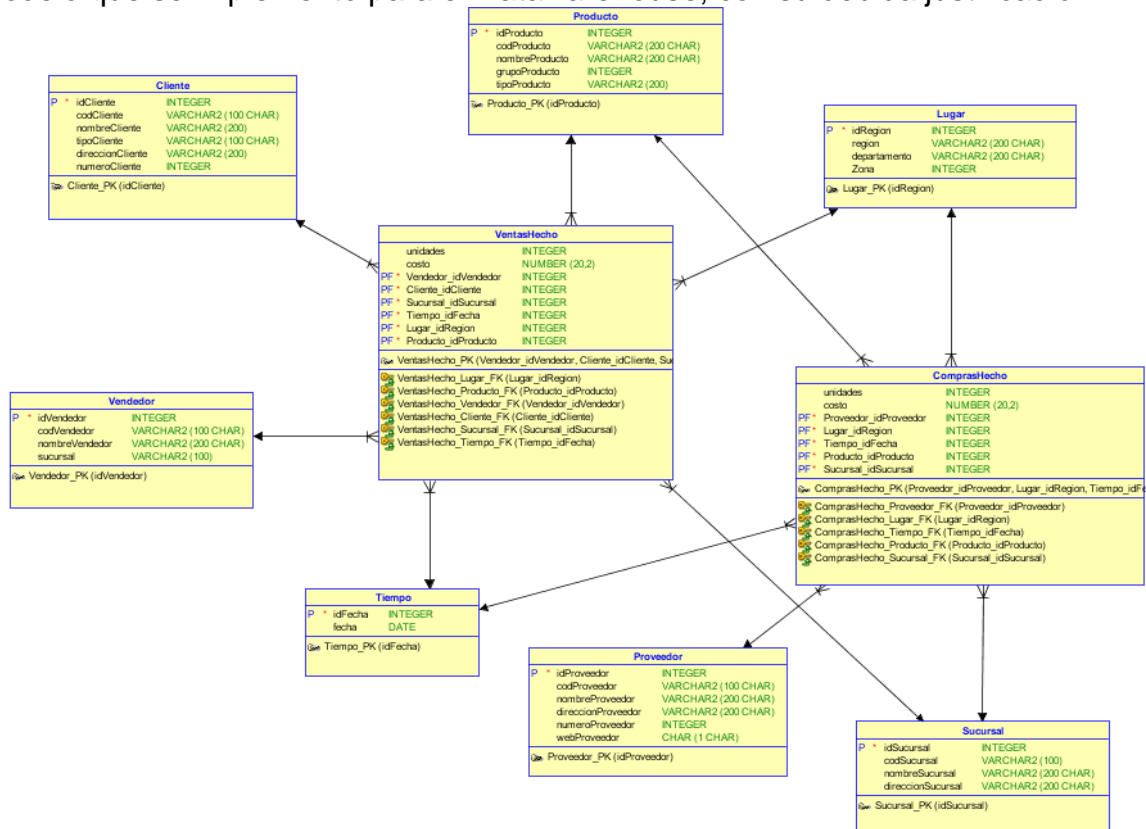
- Foreach Loop Container: utilizado para la lectura de archivos
- Data Flow Task: se utiliza para cargar la data del archivo leído en la tabla temporal ventas.
- Llenar Compras: es el proceso de llenado de tablas que van a formar luego la tabla hechos ventas.
- Llenar Hechos: es el proceso que se utiliza para llenar la tabla de hecho compras con el proceso descrito anteriormente.

Vista del proceso completo utilizado:

Sharolin Lacunza Gonzalez
201408486



Modelo que se implementó para el Datawarehouse, con su debida justificación:



El modelo fue implementado con un modelo estrella ya que las tablas centrales o las tablas de hechos comparten dimensiones y así al actualizar una dimensión se ve reflejado en todas las tablas de hechos que comparten la misma dimensión, esto es una facilidad y ventaja para cuando se generen los reportes. Es decir, su estructura hace las gestiones posteriores que se harán para analizar la data sean más simples tanto de actualizar como de modificar, incluso para insertar más registros en cada tabla y ya esté disponible para todas la tablas que las deseen. Cada dimensión cuentas con sus respectivos niveles para darle más detalle

El modelo implementado es un modelo estrella el cual cuenta con las siguientes tablas:

Dimensiones:

- Producto
 - idProducto
 - codProducto
 - nombreProducto
 - grupoProducto

- tipoProducto
- Cliente
 - idCliente
 - codCliente
 - nombreCliente
 - tipoCliente
 - direccionCliente
 - numeroCliente
- Vendedor
 - idVendedor
 - codVendedor
 - nombreVendedor
 - sucursal
- Tiempo
 - idFecha
 - fecha
- Proveedor
 - idProveedor
 - codProveedor
 - nombreProveedor
 - direccionProveedor
 - numeroProveedor
 - webProveedor
- Sucursal
 - idSucursal
 - codSucursal
 - nombreSucursal
 - direccionSucursal
- Lugar
 - idRegion
 - region
 - departamento
 - Zona

Tablas de hecho:

- VentasHecho
 - Unidades
 - Costo
- ComprasHecho
 - Unidades
 - costo