**Instalación**

Se usa Visual Studio 2019 e instalarle el Data Tools.

Instalar la extensión para SSIS Projects desde el IDE. Instalar el exe que se descarga

Configurar en propiedades del proyecto el TargetServer Version a 2019

Se usa SQL Server 2019 Developer Edition(porque la Express no la tiene). El servidor se llama DESKTOP-7EB0ON2

Instalar ACE OLEDB 12 de 32 bits para manejo de Excel

Crear un proyecto Integration Service Project. (Ver proyecto Prueba 1)

**Fundamentos**

Proceso ETL: Extracción, transformación y carga. Se realizan mediante Dataflows

Paquete: contiene la lógica de procesamiento en SSIS

El flujo de control le ayuda a organizar el orden de ejecución de todos sus componentes

Las variables tienen nombre, ámbito, tipo, valor inicial, expresión. Se pueden usar como parámetros de entrada o de salida en cada tarea. Al configurarlas se ocupan índices para referenciarlas (0,1,2,etc)

Los puntos de interrupción nos permiten ver el valor de una variable en eventos como preexecute y postexecute. Ejecutar el paquete y agregar inspección rápida o ver la pestaña de Inspeccion y poner el nombre de la variable

Las variables se pueden debuggear en Inspección Rápida cuando se esta en modo de Ejecución

El visor de datos permite visualizar los datos a través del flujo. Se activa con click derecho en una rama

**DATAWAREHOUSE-DATAMART**

Un data warehouse(OLAP) es un repositorio de varias bases de datos(OLTP). Esta conformada por dimensiones y tabla de hechos. Son mas eficientes para la obtención de datos

En el modelo de estrella hay una tabla por cada dimensión(sencillo, rápido y más utilizado)

En el modelo de copo de nieve hay varias tablas por cada dimensión

El Datamart es solo un subconjunto de un datawarehouse corporativo y se centra en solo un tema

Una arquitectura de DataMart seria la siguiente:

BD productiva-> Load(historicos) -> Stage(Solo los últimos ID)->DataMart(Productiva)

Y se agrega una BD Metadata contiene la información del usuario que modifico, cuando y las filas insertadas

**CONTROL FLOW**

**Tareas**

**Ejecutar SQL**: Ejecuta SQL y permite obtener datos de la BD

**Flujo de Datos**: Mueve datos entre orígenes y destinos, permite al usuario transformar, limpiar y modificar datos a medida que se mueven

**Inserción Masiva**: Permite volcado de datos pero sin hacer transformaciones intermedias. Se puede hacer con un BULK INSERT de SQL o por una tarea

**Ejecutar Paquete**: Permite ejecutar otros paquetes que están dentro del proyecto o que están en el servidor SQL Server o en una ubicación física externa

**Containers**

**For Loop**: Define un bucle que se realiza hasta que la condición sea false. Se definen las 3 secciones de un for considerando que las variables llevan @ al principio

**Foreach:** Permite recorrer todos los archivos de una carpeta y asociarlo a una variable. Es posible que se requiera que la conexión del archivo Excel tenga que ser dinámica. Se hace en propiedades de conexion->expressions y se asigna la variable dinámica

**Secuencial**: Es un agrupador lógico para mayor claridad del proceso

**Controlador de eventos:** En la pestaña de controladores de eventos se pueden definir acciones con tareas al ocurrir eventos como OnError, OnPreExecute, OnPostExecute, etc.

**DATA FLOW**

**Transformaciones**

**Agregado**: Aplica funciones de agregado como SUM, AVERAGE, COUNT, COUNT DISTINCT, MINIMUM, MAXIMUM, GROUP BY a una columna y copia los resultados a la salida de la transformación.

**Multicast**: Permite crear copias logicas de datos lo cual es útil cuando se necesitan aplicar transformaciones a los mismos datos

**Conversión Datos**: Convierte el tipo de dato de una columna y copia los resultados a la salida de la transformación. Util para pasar de archivos a una tabla de SQL Server y evitar errores de tipo de dato. Siempre que usamos conversiones los campos de salida cambian de nombre (considerarlo)

**Ordena**r: Permite ordenar por diferentes campos. Tambien el ordenamiento se puede hacer desde el editor avanzado del componente en la pestaña propiedades de entrada y salida

**Unión de todo**: Combinar datos de multiples fuentes hacia una sola tabla de destino. No se requiere preordenar

**Mezcla/Merge**: Combinar datos de dos conjuntos ordenados hacia una sola tabla de destino. Los datos deben estar ordenados antes del merge

**Combinacion de mezcla/Merge Join**: Unir datos de dos conjuntos ordenados con INNER, LEFT o FULL join hacia una sola tabla de destino. Los campos que se tomaran como llave deben estar ordenados antes del merge join

**Columna Derivada**: Permite agregar/reemplazar columnas con cálculos de funciones integradas de SSIS

**Recuento de filas**: Obtiene el número de filas afectadas en un Flow control y esto lo puede asignar a una variable. El valor lo toma hasta que se ejecuta la inserción/eliminación/actualización

**OLE DB Command:** Se utiliza para ejecutar una instrucción SQL (insert, update, delete) para cada fila de su flujo de datos de entrada

**División condicional**: Es similar a una estructura CASE de programación

**Script**: Permite invocar código .NET y aplicarla a cada fila de datos. Si se usa como fuente el componente permite varias salidas, si se usa como transformación admite una entrada y varias salidas, si se usa como destino admite una entrada. Se puede usar para usar campo de comprobación de cambios mediante algoritmo MD5

**PUBLICACION**

Para desplegar paquetes primero se tiene que configurar el Integration Service Catalog en el Management Studio.

Para desplegar el paquete dar click derecho y dar click en Desplegar. Y aparecerá en el Service Catalog. Se puede ejecutar ahí o desde un job

Para ejecutar los jobs validar que el SQL Server agent este activo