# SQL SERVER INTEGRATION SERVICES

# ¿Qué es SSIS?

Es una plataforma para la construcción de soluciones de integración de datos, incluyendo Extracción, Transformación y Carga (ETL) de paquetes para un Data Warehousing.

### SSIS incluye las siguientes herramientas y Wizards:

- Diseñadores: Diseñadores integrados en SQL Server Data Tools, incluyendo al ambiente para la creación de paquetes.
- Herramientas administrativas: Provee herramientas para configuración, desarrollo y ejecución de paquetes.

# ¿Qué es SSIS?

- Utilerías para línea de comando: Incluye utilerías para ejecución de paquetes desde línea de comandos
- Wizards: Incluye Wizards para la creación de tareas comunes.

# ¿Qué es SSIS?

### Conceptos en SSIS

- Paquetes
- Flujo de Control
- Flujo de Datos
- Variables
- Manejador de Errores

# Conceptos

- Paquete: Unidad de trabajo que es recuperada, ejecutada y almacenada. Son almacenados en la BD msdb o en un archivo XML con el sufijo «dtsx»
- Flujo de Control: Contiene tareas y contenedores.
   Realizan operaciones tales como preparar y copiar datos.
   Los contenedores son utilizados para agrupar tareas como una unidad lógica.
- Flujo de Datos: Contiene orígenes, transformaciones y destinos.
- Variables: Utilizados para realizar operaciones dinámicas, tales como actualización de valores de columnas.

# CREACIÓN DE PAQUETES EN SSIS

# Desarrollo de Paquetes en SSIS

### Paquete:

- Objeto CORE en SSIS.
- Contiene la lógica del negocio para manejar flujos de trabajo y procesamiento de datos.
- Permite transferir datos de orígenes a destinos y manejar la precedencia en el tiempo de cuando los datos serán procesados.

## Creación de Paquetes SSIS y Data Sources

Formas para crear paquetes:

- Utilizando en SQL Server 2014 la herramienta Wizard para Importar y Exportar (SSMS- SQL Server Management Studio).
- Crear el paquete dentro de un proyecto de SSIS (SQL Server Integration Services) en (SQL Server Data Tools)

- 1) Utilizando Wizard Exportar e Importar:
  - Permite copiar datos sin tener que crear un proyecto de SSIS.
  - Genera un paquete de SSIS que se puede ejecutar de forma inmediata o guardarlo para posteriormente modificarlo o calendarizarlo.
  - Permite traer a tablas de SQL Server datos contenido en fuentes accesibles:
    - Otras Bases de datos de SQL Server
    - Archivos planos
    - Hojas de Excel
    - Bases de datos de Access, Oracle

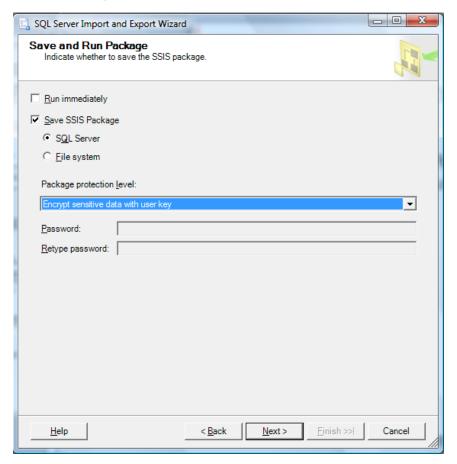
### Pasos:

- 1. Entrar a SQL Server Management Studio
- Conectarse a la instancia PRINCIPAL de SQL Server 2014 Database Engine
- 3. Desplegar todas las Bases de Datos existentes.
- Seleccionar la BD AdventureWorksDW2014 y exportar hacia la Base de Datos AdventureWorks2014:

Nota: Seleccionar las siguientes tablas a Exportar:

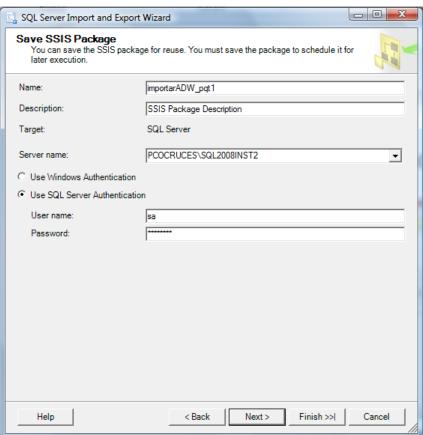
- DimAccount
- DimCustomer
- DimProduct
- DimPromotion

5. En la ventana SAVE AND RUN PACKAGE seleccionar las siguientes opciones:



Realizó: M. Lic. Octavio Cruces M.

6. Guardar el paquete con el nombre «impADW\_pqt1».



Realizó: M. Lic. Octavio Cruces M.

- El wizard para importar o exportar provee una forma rápida de mover datos de un BD origen a una Destino.
- Limitantes:
  - Solo se puede especificar un origen y un destino
  - La precedencia avanzada en el workflow no esta disponible mediante el wizard
  - El wizard no comparte orígenes de datos con otros paquetes.

### Pasos:

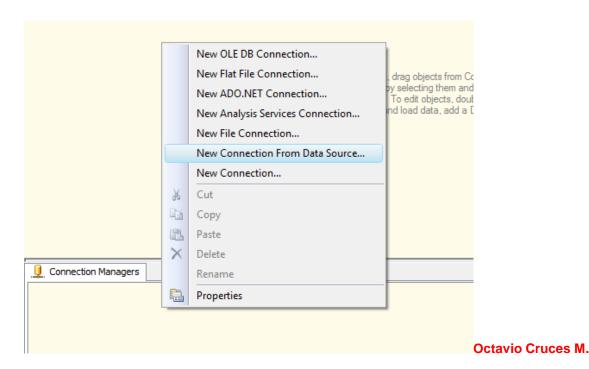
- Entrar a SQL Server Data Tools.
- 2. Crear un nuevo proyecto.
  - En Installed Templates seleccionar:
    - Integration Services
  - En la ventana derecha seleccionar :
    - Integration Services Project
  - Asignar como nombre del proyecto: ISP\_PROY1

Administrador de conexiones de paquetes

- a. Conexión con referencia al Data Source del Proyecto.
- b. Conexión independiente del Data Source del proyecto.

### Administrador de conexiones de paquetes

- a. Conexión independiente del data Source del proyecto
  - Presione Botón derecho del mouse sobre la sección «Connection Manager» y seleccione «New Connection from Data Source»



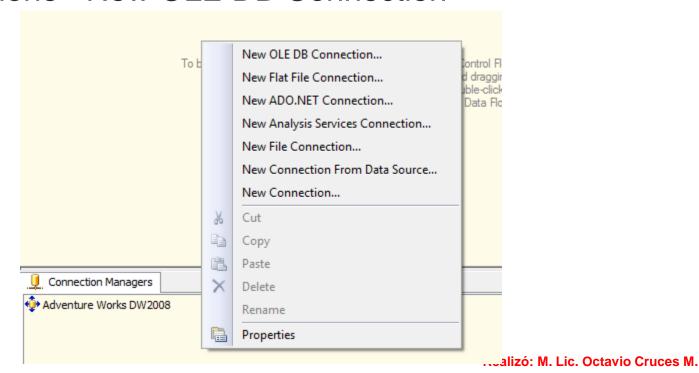
Administrador de conexiones de paquetes

b. Conexión independiente del data Source del proyecto.

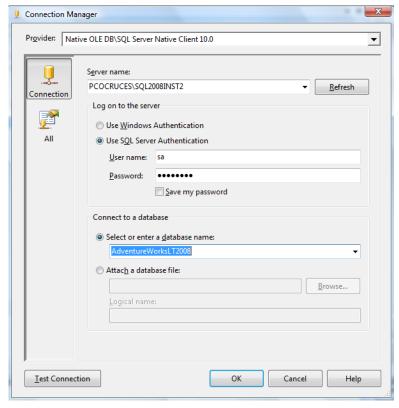
Ejemplo: Creación de conexión ODBC hacia la instancia principal a la Base de datos AdventureWorks2014

### Pasos.

- Presione botón derecho del mouse en el área «Connection Manager».
- Selectione «New OLE DB Connection»



- 1. Presione botón «New» para crear un nuevo Data Connection.
- 2. Configure los datos de la conexión



Realizó: M. Lic. Octavio Cruces M.

- 1. En SQL Server Integration Services, crear un nuevo proyecto.
- Nombrar al proyecto SSIS\_EJ1 y guardarlo en la ruta C:\SQL2014\SSIS\Ejercicios (Crear la ruta)
- 3. Renombrar el paquete generado de forma automática con el nombre Paquete1.dtsx.
- 4. Crear un nuevo paquete (Package1.dtsx)
- 5. Renombrar el paquete como DimCustomers.dtsx.
- 6. Crear un nuevo paquete con el nombre DimPromotion.dtsx.
- 7. Crear dos Connection Manager con las siguientes propiedades:

Connection Manager 1	
Name	Adventure Works DW2014
Ámbito	Global
Provider	Native OLE DB\SQL Server Native Client 11
Server Name	(Nombre_PC)
Authentication	SQL Server
Select Or Enter A Database Name	AdventureWorksDW2014

Connection Manager 2	
Name	Adventure Works 2014
Ámbito	Local para paquete (Paquete1.dtsx)
Provider	Native OLE DB\SQL Server Native Client 10
Authentication	SQL Server
Select Or Enter A Database Name	AdventureWorks2014

Connection Manager 3	
Name	Adventure Works 2014
Ámbito	Local para paquete: DimCustomer.dtsx
Provider	Native OLE DB\SQL Server Native Client 10
Authentication	SQL Server
Select Or Enter A Database Name	AdventureWorks2014

Connection Manager 3	
Name	Adventure Works 2014
Ámbito	Local para paquete: DimPromotion.dtsx
Provider	Native OLE DB\SQL Server Native Client 10
Authentication	SQL Server
Select Or Enter A Database Name	AdventureWorks2014

# SSIS – Creación y Edición de Objetos de Flujo de Control

- Un paquete de flujo de control es el proceso para controlar la ruta de ejecución de un paquete.
- Permite poner condiciones para controlar la secuencia de tareas basadas en el éxito o el fracaso de una tarea anterior o sobre los resultados de las expresiones.
- Permite agrupar tareas dependiendo los propósitos de la organización.

### Contenedor:

Proporcionan estructura en paquetes y servicios a las tareas de flujo de control

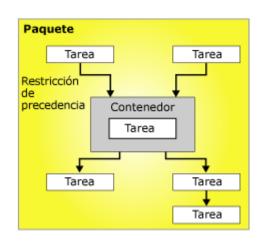
### Tareas:

Realizan el trabajo en los paquetes

### Restricciones de Precedencia

Conectan contenedores y tareas de paquetes en un flujo de control ordenado.

Controlan la secuencia para las tareas y contenedores.



Tarea	Descripción
ActiveX Script Task	Ejecuta código de VBScript y Jscript. Útil en migración de paquetes DTS a SSIS
Analysis Services Execute DDL Task	Ejecuta código XML para análisis (XMLA) frente a una Base de datos SSAS.
Analysis Services Processing Task	Permite el procesamiento de objetos de SSAS a través de paquetes de SSIS.
Bulk Insert Task	Operaciones de Copia de datos (SSMS)
Data Flow Task	Permite procesamiento de datos entre Orígenes y destinos
Data Mining Query Task	Realiza consultas de minería de datos y permite capturar los resultados para su análisis

Tarea	Descripción
Data Profiling Task	Permite el análisis de datos fuente para patrones, llaves candidatas y estadísticas.
Execute DTS 2000 Package Task	Ejecuta paquetes DTS sin SSIS
Execute Package Task	Ejecuta otros paquetes de SSIS
Execute Process Task	Ejecuta operaciones de línea de comando tales como un programa o un archivo batch
Execute SQL Task	Ejecuta código SQL
File System Task	Permite copiar, mover y borrar archivos

Tarea	Descripción
FTP Task	Envía y recibe archivos entre los archivos de sistema y un servidor FTP
Script Task	Ejecuta Microsoft Visual Basic 2008 o Microsoft Visual C# 2008 con un paquete de SSIS
Send Mail Task	Envían mensajes de correo a través de un servidor SMTP.
Transfer Database Task	Tarea que copia Objetos SQL Server de un sistema a otro, incluyendo Bases de datos, Agentes Job de SQL Server, Mensajes de error, logins, etc.
Web Service Task	Permite conectar a un Web Service para enviar y recibir información.
WMI Data Reader Task	Permite ejecutar una consulta Windows Management Instrumentation (WMI) para capturar información del servidor.

Tarea	Descripción
WMI Event Watcher Task	Espera por la ejecución de un evento particular
XML Task	Consultas y distingue entre múltiples archivos XML en el servidor.

### Tareas de Planes de Mantenimiento

Tarea	Descripción
Back Up Database Task	Respalda una o más Bases de Datos
Check Database Integrity Task	Verifica la integridad de la Base de Datos
Execute SQL Server Agent Job Task  Execute	Ejecuta uno o más jobs en una instancia de SQL Server
T-SQL Statement Task  History Cleanup	Ejecuta comandos de Transact SQL Server (Similar a la tarea Execute SQL Task) Elimina historial de las tablas en la base de datos msdb

Tarea	Descripción
Maintenanc e Cleanup Task	Borra archivos o reportes de texto de planes de mantenimiento.
Notify Operator Task	Envía notificaciones a operadores (SQL Server Agent)
Rebuild Index Task	Reconstruye índices
Reorganize Index Task	Reorganiza índices
Shrink Database Task	Reducción del tamaño de log y data
Update Statistics Task	Actualización de estadísticas

# SSIS – Tarea de Flujo de Datos

- Es una tarea especial que encapsula la extracción, transformación y carga de operaciones en SSIS.
- Aparece en la página de flujo de control como una tarea normal pero tiene su propia opción (pestaña) ya que consiste de varios componentes:
  - Source:
    - Define una conexión a un almacenamiento de datos relacional o norelacional tales como SQL Server o una hoja de cálculo de Excel.
    - Tiene una salida con múltiples columnas.

# SSIS – Tarea de Flujo de Datos

### Transformation:

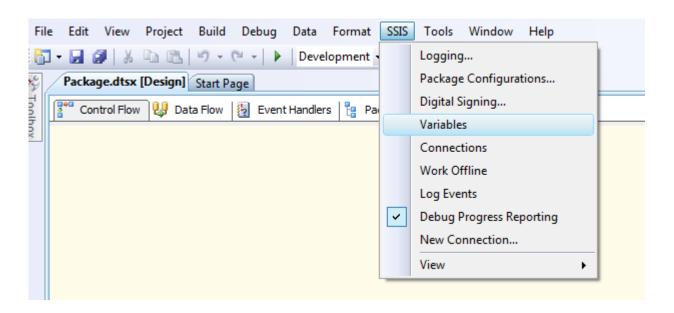
- Transforma los datos mapeando el destino, limpiando, agregando y modificando datos y esquemas.
- Una transformación puede ser una sola columna entre un origen y un destino, pero puede ejecutar complejas modificaciones y operaciones de limpieza de datos.

### Destination:

- Identifica una conexión a un almacenamiento de salida en un dataset en memoria.
- Tiene al menos una entrada la cual al menos tiene una columna.

# SSIS – Variables de Paquetes

Las variables dentro de SSIS son una forma de integrar objetos pasando valores entre tareas y contenedores, aceptando valores de fuentes externas o construyendo código dinámico cuando es ejecutado.



# SSIS – Variables de Paquetes

### Tipos de variables:

### 1. Variables de Sistema:

- ✓ Son utilizadas por los paquetes para rastrear metadatos como el nombre del paquete y el usuario que ejecuta el paquete.
- ✓ No son editables.
- ✓ Pueden ser referenciadas dentro de tareas y contenedores

### Variables de Usuario:

✓ Creadas y definidas por un usuario para algún propósito en el paquete.

Crear una tarea de Flujo de Control y examinar su ejecución (Debug).

- 1. Abrir el proyecto SSIS\_EJ1 generado en el ejercicio 1
- Editar el paquete Paquete1.dtsx (Botón derecho del mouse open)
- 3. Agregar a la ventana de diseño del paquete la tarea «EXECUTE SQL TASK» de la caja de herramientas.
- 4. Crear un Connection Manager hacia la BD AdventureWorks2014.
- Editar el objeto «Execute SQL Task» presionando doble click o botón derecho del mouse y posteriormente Edit.
- 6. Cambie la propiedad conexión para usar la conexión AdventureWorks2014

6. En la propiedad «SQLStatement» escriba el siguiente código.

**UPDATE** Production.Product

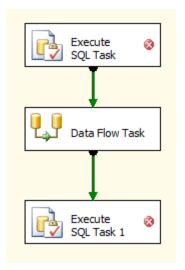
SET ProductLine = 'S'

WHERE ProductLine IS NULL

- Seleccione OK en la caja de dialogo para regresar al diseñador de SSIS. Presione botón derecho sobre la tarea «Execute SQL Task» y renómbrela como «Actualizar ProductLine»
- 8. Seleccione «Save All» de la barra de herramientas de Data Tools
- Ejecute el proyecto.
- 10. Seleccione la pestaña "Execution Results" ó "Progress" para visualizar el detalle de la ejecución.

Modificar el paquete de flujo de control ETL DimCustomer.

- 1. Seleccione el paquete DimCustomer.
- Seleccione de la caja de herramientas dos objetos «Execute SQL Task» y un «Data flow Task»
- Conecte el primer objeto «Execute SQL Task» a una tarea «Data Flow Task» y conecte la restricción de precedencia del «Data Flow Task» hacia la segunda «Execute SQL Task» (Ver imagen en siguiente diapositiva)



4. Renombre el primer «Execute SQL Task» a «Truncate Update Table», renombre el segundo «Execute SQL Task» como «Batch Updates».

5. Entrar a SQL Server Management Studio, conectarse a Database Engine y ejecutar lo siguiente en BD AdventureWorksDW201x.

```
CREATE TABLE stgDimCustomerUpdates(
[CustomerKey] [int] NOT NULL,
[GeographyKey] [int] NULL,
[CustomerAlternateKey] [nvarchar](15) NOT NULL,
[Title] [nvarchar](8) NULL,
[FirstName] [nvarchar](50) NULL,
[MiddleName] [nvarchar](50) NULL,
[LastName] [nvarchar](50) NULL,
[NameStyle] [bit] NULL,
[BirthDate] [date] NULL,
[MaritalStatus] [nchar](1) NULL,
[Suffix] [nvarchar](10) NULL,
[Gender] [nvarchar](1) NULL,
[EmailAddress] [nvarchar](50) NULL,
                                                 Realizó: M. Lic. Octavio Cruces M.
```

```
[YearlyIncome] [money] NULL,
[TotalChildren] [tinyint] NULL,
[NumberChildrenAtHome] [tinyint] NULL,
[EnglishEducation] [nvarchar](40) NULL,
[SpanishEducation] [nvarchar](40) NULL,
[FrenchEducation] [nvarchar](40) NULL,
[EnglishOccupation] [nvarchar](100) NULL,
[SpanishOccupation] [nvarchar](100) NULL,
[FrenchOccupation] [nvarchar](100) NULL,
[HouseOwnerFlag] [nchar](1) NULL,
[NumberCarsOwned] [tinyint] NULL,
[AddressLine1] [nvarchar](120) NULL,
[AddressLine2] [nvarchar](120) NULL,
[Phone] [nvarchar](20) NULL,
[DateFirstPurchase] [date] NULL,
[CommuteDistance] [nvarchar](15) NULL
```

- 6. Después de crear la tabla regrese al paquete DimCustomer en SSIS y edite la tarea «Truncate Update Table».
- 7. Cambie la propiedad de conexión a AdventureWorks DW201X.
- 8. En la propiedad SQLStatement escriba el siguiente código.

TRUNCATE TABLE dbo.stgDimCustomerUpdates

 Editar la tarea «Batch Updates» y configure la propiedad de conexión hacia AdventureWorks DW201X.

10. En la propiedad SQLStatment capture el siguiente comando UPDATE:

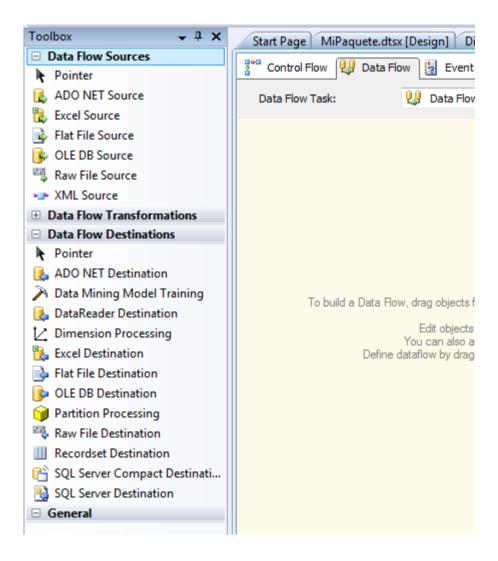
INSERT INTO stgDimCustomerUpdates
SELECT \* FROM DimCustomer

11. Presione OK guarde el paquete.

### SSIS – Flujo de Datos

- Una tarea puede tener cero, uno o más flujos de datos.
- El flujo de datos puede tener tres tipos de objetos en la caja de herramientas:
  - Data Flow Sources Adapters
  - Data Flow Transformations
  - Data Flow Destination Adapters

## SSIS – Flujo de Datos



### SSIS – Data Flow Source Adapters

- Data Flow Source Adapters utiliza conexiones de paquetes, las cuales apunta a la instancia del servidor o a la dirección donde se encuentra el archivo del Data Source.
- Un Source Adapter extrae los datos de un origen y los mueve hacia el flujo de datos, donde serán modificados e enviados a un destino.

## SSIS – Data Flow Source Adapters

Data Flow Source	Propósito
ADO.NET Source	Provee conexiones a tablas o consultas a través del proveedor ADO.NET
Excel Source	Permite la extracción de una hoja de calculo definir en un archivo de excel.
Flat File Source	Conexión a un archivos planos de diferentes código de páginas
OLE DB Source	Conexión a proveedores instalados de OLE DB tales como SQL Server, Access, Oracle.
Raw File Source	Almacena datos nativos de SSIS en archivo binario útiles para organizar datos.
XML Source	Permite extraer datos de archivos XML. Requiere un schema XML para definir asociaciones de datos.

#### SSIS – Data Flow Destinations

 Al igual que Data Flow Sources, utiliza conexiones de paquetes, sin embargo Data Flow Destinations son endpoints en un paquete, definiendo la ubicación en la cual los datos deberían ser colocados.

#### Ejemplo:

Envío de datos a un archivo de Excel de una tabla de una Base de datos, el destino sería el adaptador de destino de Excel.

### SSIS – Data Flow Destinations

Data Flow Destination	Propósito
ADO.NET Destination	Permite la inserción de datos utilizando el manejador de ADO.NET
Data Mining Model Training	Permite transferir datos del Data Flow hacia un modelo de datamining en SSAS
DataReader Destination	Permite poner datos en un recordset de ADO.NET que puede ser referenciado por programación
Dimension Processing	Permite que un SSAS Dimensional sea procesado directamente del flujo de datos
Excel Destination	Utilizado para la inserción de datos a Excel
Flat File Destination	Permite inserción de datos a archivos planos
OLE DB Destination	Utiliza OLE-DB para insertar registros

Realizó: M. Lic. Octavio Cruces M.

### SSIS – Data Flow Destinations

Data Flow Destination	Propósito
Partition Processing	Permite particiones SSAS sean procesadas directamente del Data Flow
Raw File Destination	Almacena datos SSIS Nativos en un archivo binario
Recordset Destination	Toma datos del flujo de datos y crea un recordset en una variable de paquete de tipo objeto.
SQL Server Compact Destination	Permite enviar datos a dispositivos móviles corriendo SQL Server Mobile
SQL Server Destination	Provee un destino especifico de alta velocidad a SQL Server 2008 si el paquete esta corriendo bajo SQL Server

- Da la capacidad para modificar y manipular datos en el flujo de datos.
- La transformación permite realizar operaciones sobre un registro o sobre múltiples registros a la vez.
- Las transformaciones se dividen en las siguientes categorías:
  - Logical Row-Level Transformations
  - Multi- input o Multi-Output Transformations
  - Multi-Row Transformations
  - Advanced Data-Preparation Transformations

Logical Row-Level Transformations

Es la más sencilla y común operación de transformación en registros. Esta transformación trabaja a nivel registro.

#### Usos:

- Cálculos matemáticos
- Conversión un valor de texto a valor numérico o decimal
- Reemplazar Nulos con otros valores

Multi- input o Multi-Output Transaformations

Permite trabajar con más de un dato de entrada o puede generar más que una sálida respectivamente. Provee la capacidad para combinar datos y dar al DATA FLOW la capacidad para procesar de una o más fuentes hacia uno o más destinos.

Multi-Row Transformations

Transformaciones que trabajan bajo criterios de múltiples entradas de registros o generan múltiples registros de salida de un registro único de entrada.

- Asociar registros basados en grupos definidos y agregaciones generadas tales como SUM, MAX, MIN, COUNT
- Ordenar entradas basadas en columnas de orden definidas y la dirección de ordenamiento.

Advanced Data Preparation Transformations

Permite generar operaciones avanzadas en registros a lo largo del Data Flow.

- Aplicar registros de entrada a un modelo de minería de datos
- Procesar cambios en dimensiones de cubos
- Analizar entrada de columnas de texto contra un conjunto de palabras definidas por el usuario

- 1. Abrir proyecto SSIS\_EJ1. Seleccionar el paquete Paquete1.dtsx.
- 2. En el Flujo de control crear un contenedor de Secuencias
- Dentro del contenedor agregar la tarea: "Actualizar Product Line".
- Dentro del contenedor agregar una tarea: «Data Flow Task» como salida de la tarea del paso 3.
- 5. Seleccionar la pestaña Data Flow en Paquete1.dtsx

- 6. De la caja de Herramientas arrastrar «OLE DB Source» colocado debajo del grupo de «Other Sources» hacia el área de trabajo del Data Flow. Presionar botón derecho del mouse sobre el elemento OLE DB Source y seleccionar EDIT.
- Seleccione «AdventureWorks201X» en «OLE DB connection manager», en la opción «Data access mode» seleccione «SQL command»
- 8. En el cuadro de texto de SQL Command escriba la siguiente consulta:

SELECT SH.OrderDate, SD.LineTotal, P.ProductLine

FROM Sales.SalesOrderHeader SH

INNER JOIN Sales. Sales Order Detail SD

ON SH.SalesOrderID = SD.SalesOrderID

INNER JOIN Production Product P

ON SD.ProductID = P.ProductID

- Seleccione la opción «Columns» del recuadro de la izquierda y valide que las columnas OrderDate, LineTotal y ProductLine son mostradas como columnas disponibles en el adaptador origen.
- 10. Presione OK de la caja de dialogo del OLE DB Source Editor.
- 11. De la caja de herramientas seleccione y arrastre hacia el área de trabajo «Aggregate» debajo del objeto «Common».

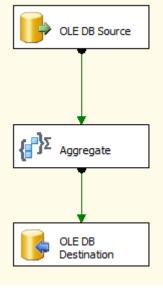
- 12. Ligue la salida de «OLE DB Source» con el objeto «Aggregate».
- 13. Edite el objeto «Aggregate» (doble click sobre el objeto):
  - a. Seleccione OrderDate de la lista de columna de entrada y verifique que la operación por default «Group by» esta seleccionada para el nuevo registro.
  - b. Agregue una segunda columna de entrada seleccionando la columna LineTotal de la lista. Para la columna operación seleccione SUM y en alias escriba SubTotal.
  - Agregue una tercer columna seleccionando la columna ProductLine.

- d. Verifique que la operación por default es «Group by»
- e. Click OK en la caja de dialogo del editor de Aggregate.
- 14. En la caja de herramientas de Data Flow, navegue hacia el grupo de objetos «Other Destinations», seleccione y arrastre el objeto «OLE DB Destination» y líguelo a la salida del objeto «Aggregate».
- 15. Presione botón derecho del mouse en el adaptador «OLE DB Destination» y seleccione Edit, modifique las siguientes opciones:
  - Verifique que el OLE DB Connection Manager tenga seleccionado: AdventureWorks201X.

- Seleccione el botón NEW que se encuentra del lado derecho de «Name of the table or the view».
- En la caja de dialogo cambie el nombre de la nueva tabla a Sales\_Summary como se muestra a continuación:

```
CREATE TABLE [Sales_Summary] (
[OrderDate] datetime,
[SubTotal] numeric(38,6),
[ProductLine] nvarchar(2))
```

- d. Presione OK
- e. En la opción Mappings del la caja de dialogo del editor de OLE Destination, asegurarse que todas las columnas estén mapeadas de un origen a un destino.
- f. Selección OK para guardar los cambios.



- 16. Seleccione botón derecho sobre el área de diseño de Data Flow y presione Execute. Valide que se complete con éxito la ejecución.
- 17. Presione Botón Stop en la barra de herramientas (Ctrl + F5) y guarde todos los cambios (Ctrl+Shift+S).

### Manejador de Eventos

- Permite crear flujos de trabajo para el manejo de errores, avisos, tareas completadas, contenedores o paquetes.
- Permite configurar el alcance del manejador de errores

## Manejador de Eventos

Event	When Event Is Raised
OnError	When an error occurs
OnExecStatusChanged	When an executable 's status changes
	When an informational event is raised during the validation
OnInformation	and execution of an executable
OnPostExecute	When an executable completes
OnPostValidate	When an executable 's validation is complete
OnPreExecute	Before an executable runs
OnPreValidate	Before an executable 's validation begins
OnProgress	When measurable progress has happened on an executable
OnQueryCancel	When a query has been instructed to cancel
OnTaskFailed	When a task fails
OnVariableValueChanged	When a variable is changed at runtime
OnWarning	When a warning occurs in your package

### Explorador de Objetos

- Consolida todos los paneles de diseño en una sola vista
- Lista todas las tareas, conexiones, contenedores, manejadores de evento, variables y transformaciones de los paquetes.