



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

Laboratorio de: Inteligencia de negocios

Tema: Práctica ETL

Nombre: Danny Díaz Padilla

Fecha: 23/11/2019

1. Objetivos:

1.1. Objetivo General

Utilizar SQL Data Tools para crear un procedimiento ETL.

1.2. Objetivos Específicos

Completar el tutorial 1 que ofrece Microsoft para la práctica.

Leer información de un archivo plano como entrada.

2. Marco teórico:

ETL

Consiste en un proceso en donde se extrae información de distintas fuentes, se aplica transformaciones en orden de estandarizar la información y se carga a una base de datos de destino.

Para llevar a cabo de manera correcta el proceso de extracción, primera fase de los procesos ETL, hay que seguir los siguientes pasos:

- Extraer los datos desde los sistemas de origen.
- Analizar los datos extraídos obteniendo un chequeo.
- Interpretar este chequeo para verificar que los datos extraídos cumplen la pauta o estructura que se esperaba. Si no fuese así, los datos deberían ser rechazados.
- Convertir los datos a un formato preparado para iniciar el proceso de transformación

SQL Data Tools

Es un conjunto de herramientas de datos de SQL Server (SSDT) que transforma el desarrollo de bases de datos al presentar un modelo ubicuo y declarativo que abarca todas las fases del desarrollo y el mantenimiento y la actualización de bases de datos dentro de Visual Studio. Las capacidades de diseño de Transact-SQL de SSDT se pueden usar para compilar, depurar, mantener y refactorizar bases de datos, permitiendo la integración total del motor de base de datos SQL Server con la herramienta de desarrollo Visual Studio.

SQL Server Data Tools se considera como la siguiente evolución de Visual Studio basada en herramientas de base de datos que proporciona a los desarrolladores una herramienta única para apoyar las necesidades del desarrollo de bases de datos.

La herramienta para el desarrollo de proyectos de Business Intelligence en Visual Studio 2008 y versiones anteriores la conocíamos como Business Intelligence Development Studio (BIDS), ahora para Visual Studio 2010, 2012 y 2013 estará bajo el nombre de SQL Server Data Tools, sumándose a este nuevo conjunto de herramientas para el desarrollo de base de datos.

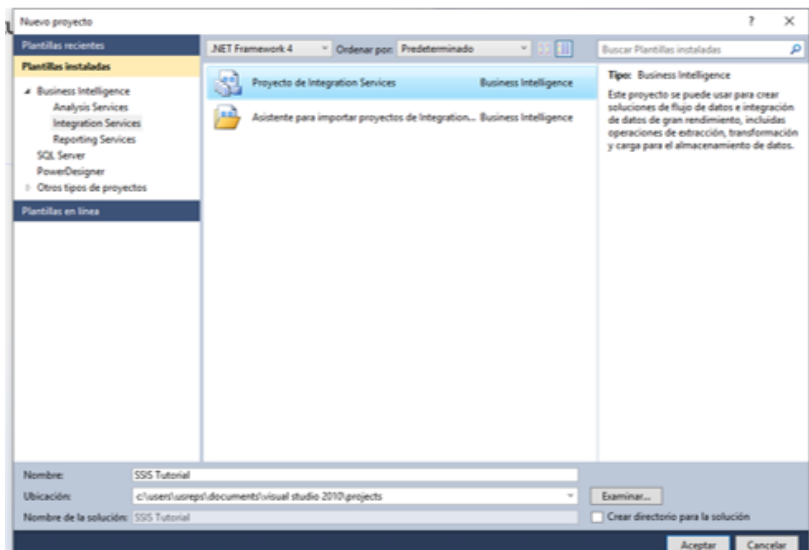




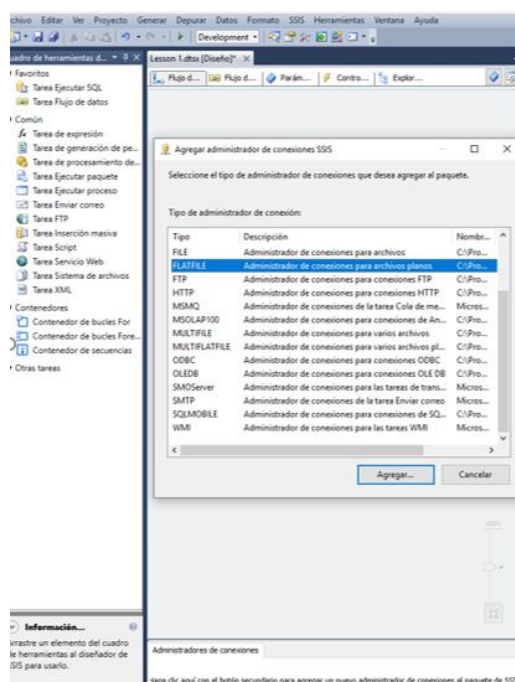
ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

3. Desarrollo de la práctica:

Paso 1: Creación de un proyecto de Integration Services



Paso 2: Adición y configuración de un administrador de conexiones de archivos planos



Editor del administrador de conexiones de archivos planos

Nombre del administrador de conexiones: Sample Flat File Source Data

Descripción:

Seleccione un archivo y especifique sus propiedades y formato.

Nombre de archivo: P:\Downloads\SampleCurrencyData.txt Examinar...

Configuración regional: Español (Ecuador) ☐ Unicode

Página de códigos: 1252 (ANSI - Latin I)

Formato: Delimitado

Calificador de texto: <ninguno>

Delimitador de filas de encabezados: CR/LF

Filas de encabezados que se omitirán: 0

☒ Nombres de columna de la primera fila de datos

Aceptar Cancelar Ayuda

Editor del administrador de conexiones de archivos planos

Nombre del administrador de conexiones: Sample Flat File Source Data

Descripción:

Configure las propiedades de cada columna.

Columna	Propiedades
1	<p>USD</p> <p>1.105000</p> <p>0.99980004</p>

Varios

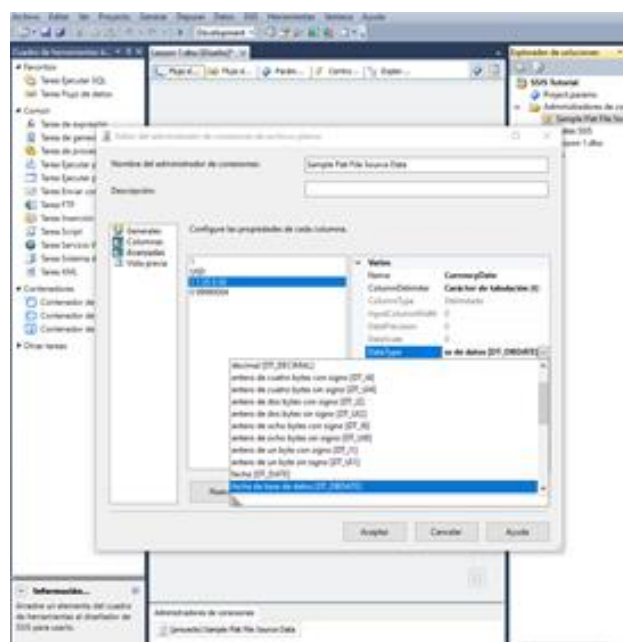
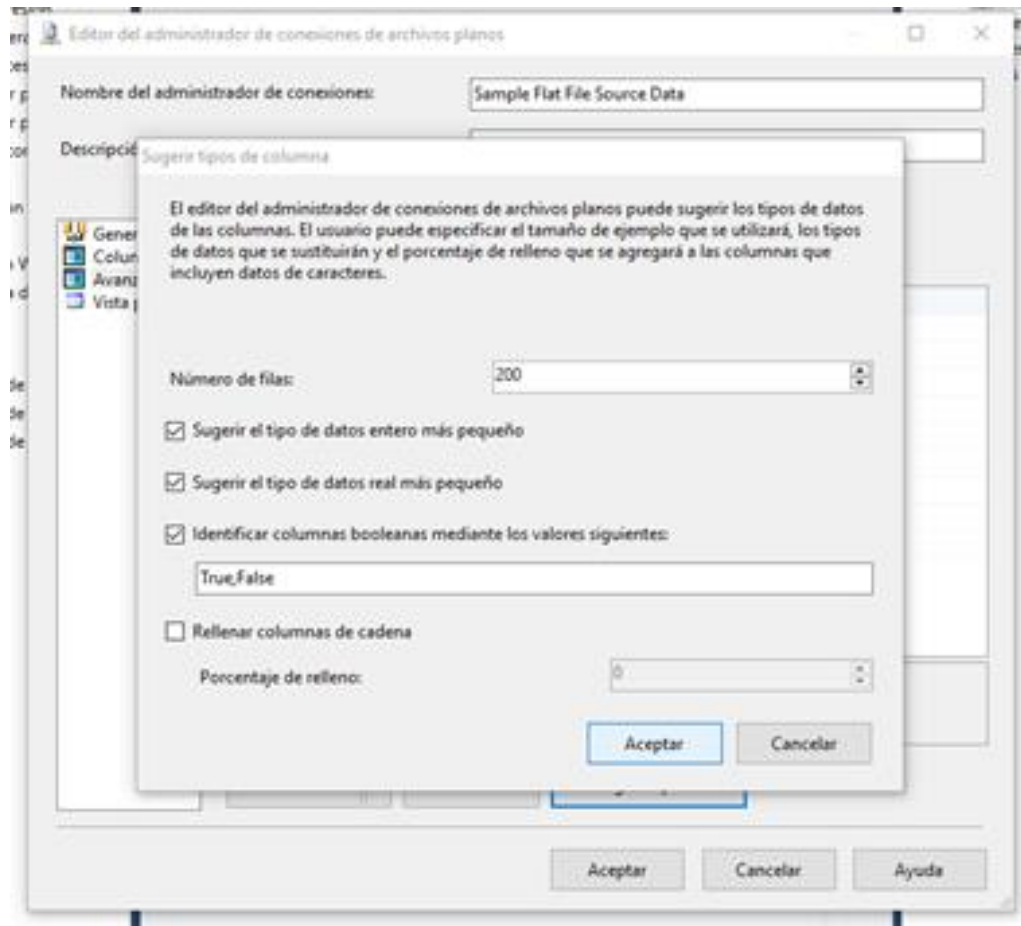
Propiedad	Valor
Name	EndOfDayRate
ColumnDelimiter	(CR)(LF)
ColumnType	Delimitado
InputColumnWidth	0
DataPrecision	0
DataScale	0
DataType	cadena [DT_STR]
OutputColumnWidth	50
TextQualified	True

Nuevo Eliminar Sugerir tipos...

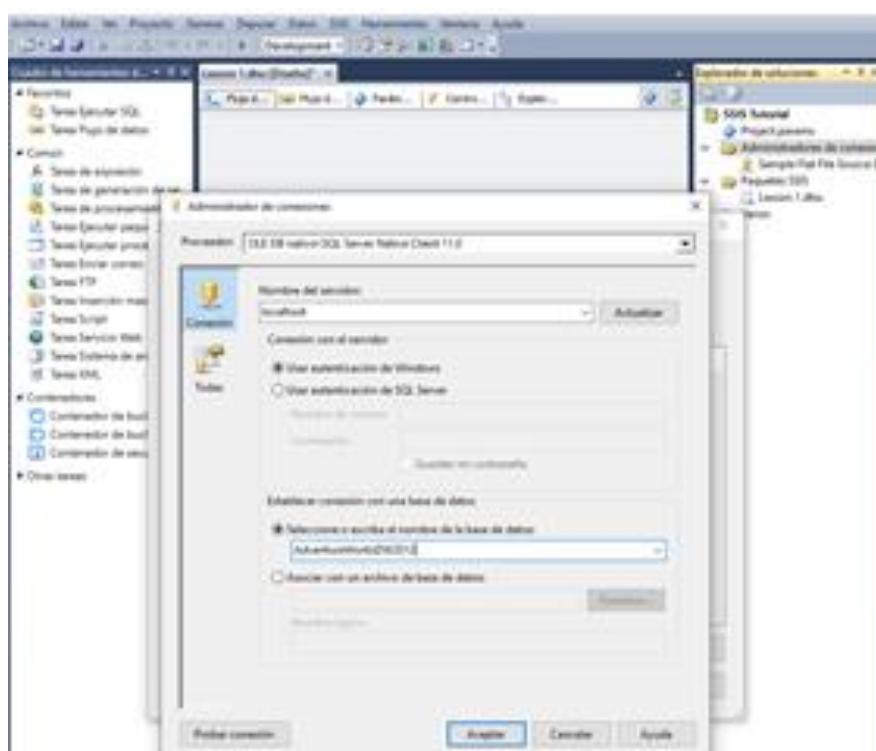
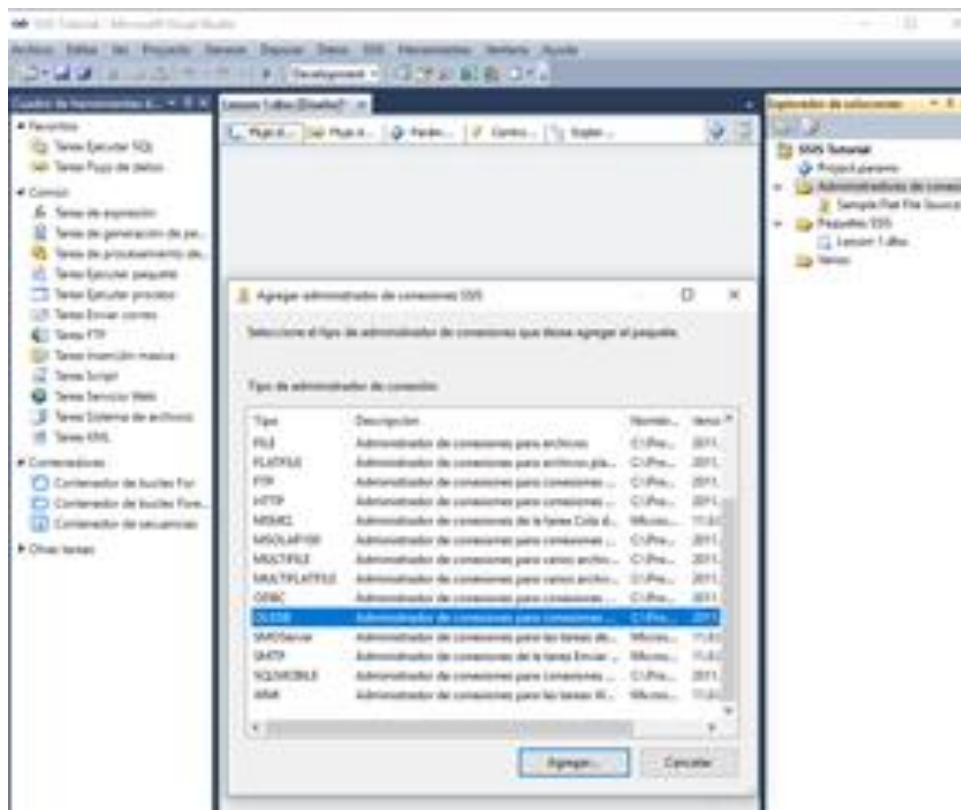
Aceptar Cancelar Ayuda



ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

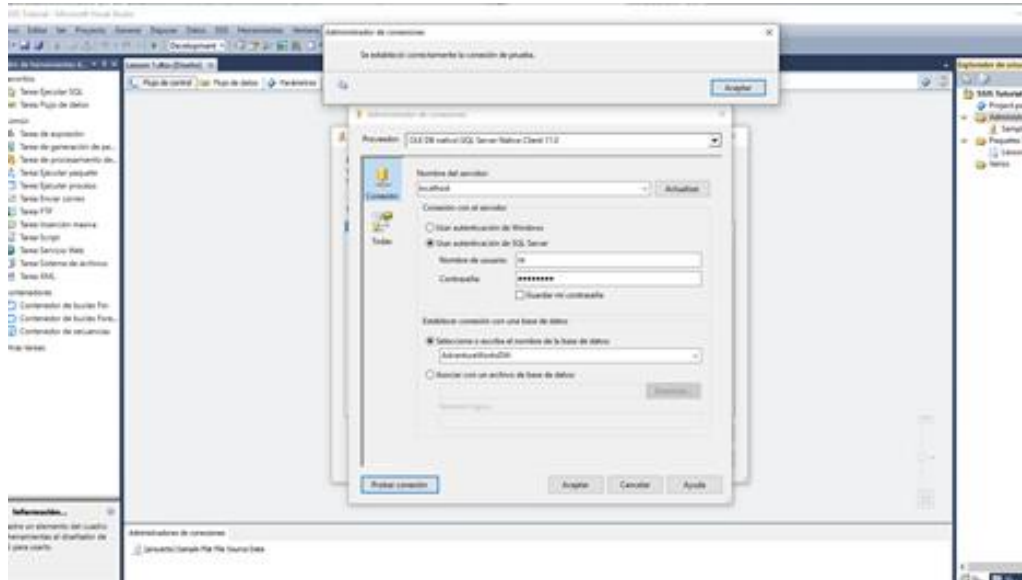


Paso 3: Adición y configuración de un administrador de conexiones OLE DB

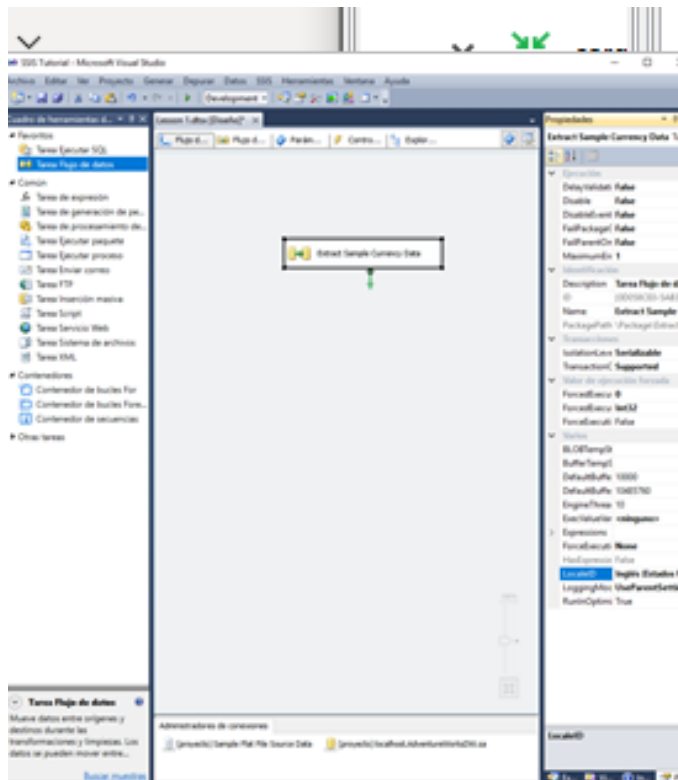




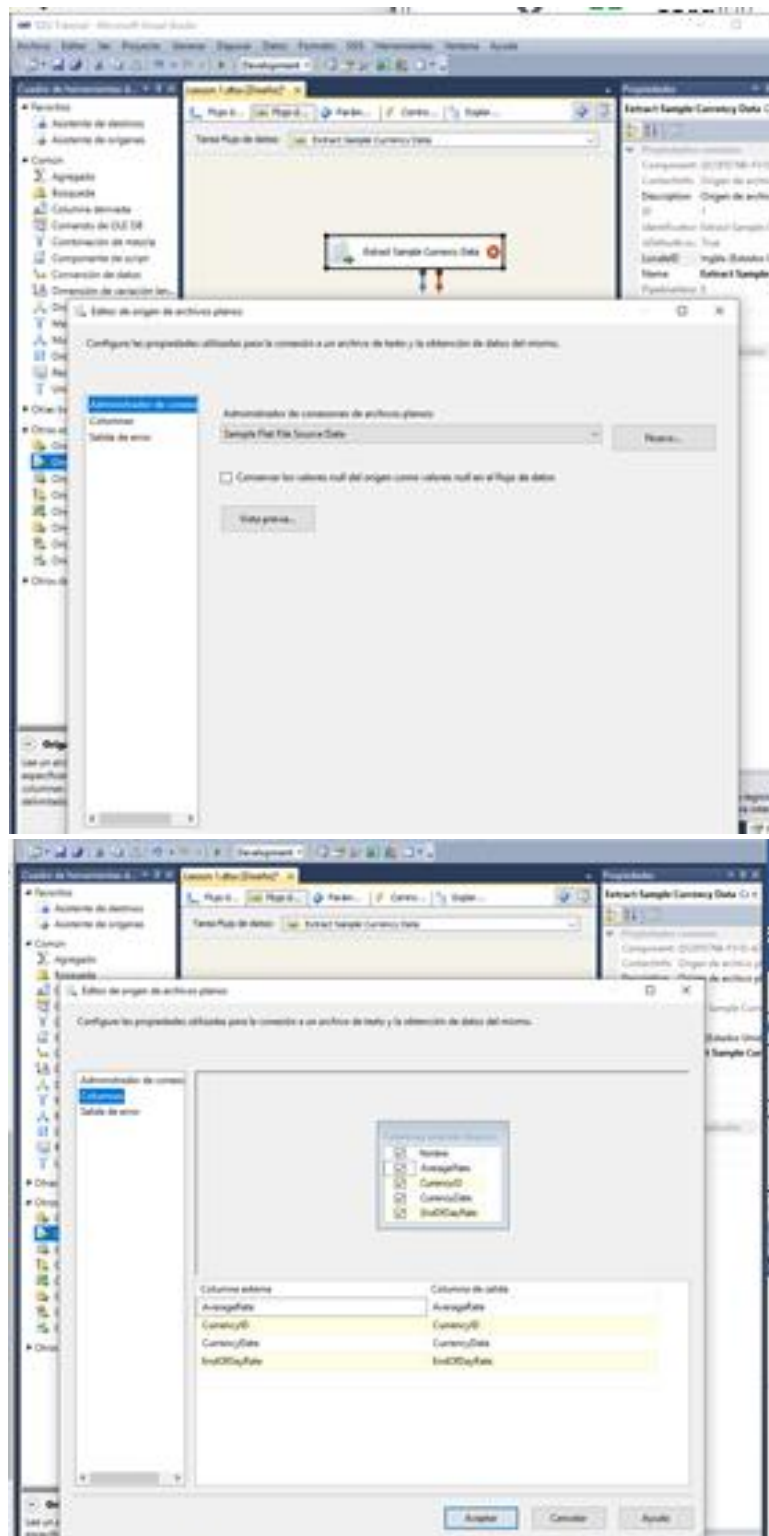
ESCUELA POLITÉCNICA NACIONAL FACULTAD DE INGENIERÍA DE SISTEMAS INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN



Paso 4: Adición de una tarea Flujo de datos al paquete



Paso 5: Adición y configuración del origen de archivo plano



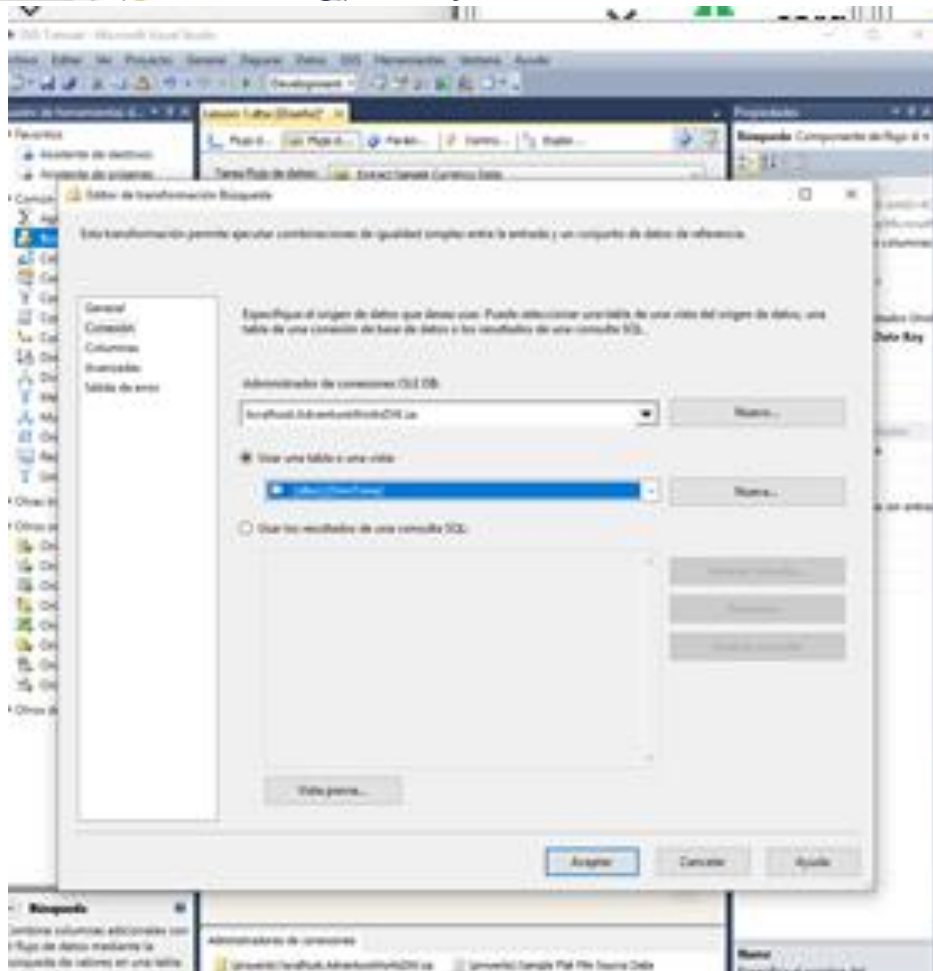


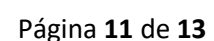
ESCUELA POLITÉCNICA NACIONAL FACULTAD DE INGENIERÍA DE SISTEMAS INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

Paso 6: Adición y configuración de transformaciones de búsqueda

The screenshots illustrate the steps for adding and configuring search transformations:

- Top Left:** The 'Edit Search Transformation' dialog box. It shows the 'Cache' mode set to 'Full cache' and the 'Type of cache' set to 'Administrator of connections OLE DB'. The 'Specify which table or file is the source of data' field is empty.
- Top Right:** The 'Edit Search Transformation' dialog box. It shows the 'Specify which table or file is the source of data' field set to 'Table of data'. The 'Administrator of connections OLE DB' is set to 'Microsoft Access Driver (*.mdb, *.accdb)'. The 'Use the results of a query (SQL)' checkbox is checked, and a list of SQL queries is shown.
- Bottom Left:** The 'Edit Search Transformation' dialog box. It shows the 'Specify which table or file is the source of data' field set to 'Table of data'. The 'Administrator of connections OLE DB' is set to 'Microsoft Access Driver (*.mdb, *.accdb)'. The 'Use the results of a query (SQL)' checkbox is checked, and a list of SQL queries is shown.
- Bottom Right:** The 'Edit Search Transformation' dialog box. It shows the 'Specify which table or file is the source of data' field set to 'Table of data'. The 'Administrator of connections OLE DB' is set to 'Microsoft Access Driver (*.mdb, *.accdb)'. The 'Use the results of a query (SQL)' checkbox is checked, and a list of SQL queries is shown.







ESCUELA POLITÉCNICA NACIONAL
FACULTAD DE INGENIERÍA DE SISTEMAS
INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN

4. Análisis de resultados:

La práctica solo funciona con la base de datos AdventureWorksDW2012, dado que si se usa otra versión los tipos de datos (sobre todo de fechas) cambian y se debe utilizar tipos de datos TimeStamp como reemplazo.

De forma gráfica se puede realizar de forma más entendible para el equipo de trabajo los procesos ETL.

5. Conclusiones y recomendaciones:

- SQL Data Tools permite de forma gráfica extraer la información y realizar junturas de tablas según necesitemos.
 - Algunos tipos de datos no son soportados por AdventureWorksDW2012. Por lo que se debe buscar tipos de datos similares o reestructurar el esquema.
 - La transformación de datos permite estandarizar la forma en que se almacenarán los datos independientemente si es plano o estructurado en una base de datos relacional.
 - Al cambiar de versión de SQL Data Tools no reconoce el archivo, de forma que el objetivo de terminar la primera lección no se pudo cumplir.
-
- Se recomienda utilizar el formato de Inglés (Estados Unidos) para realizar la práctica.
 - Es necesario utilizar credenciales SQL y no de Windows para que la práctica funcione.

6. Bibliografía:

<https://docs.microsoft.com/es-es/sql/integration-services/ssis-how-to-create-an-etl-package?view=sql-server-2014>

<https://docs.microsoft.com/es-es/sql/integration-services/lesson-1-create-a-project-and-basic-package-with-ssis?view=sql-server-2014>

<https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/qu-son-los-procesos-etl>

<https://miblogtecnico.wordpress.com/tag/sql-server-data-tools/>