



Universidad César Vallejo

Pregrado

GESTIÓN DE DATOS E INFORMACIÓN II

30
años

Licenciada por
SUNEDU
para que puedas
salir adelante

SESIÓN 09 INTEGRANDO DATOS II



SQL SERVER INTEGRATION SERVICE

- ✓ Es una plataforma para la creación de soluciones empresariales de transformaciones de datos e integración de datos. Use **Integration Services** para resolver complejos problemas empresariales mediante la copia o descarga de archivos, la carga de almacenamientos de datos, la limpieza y minería de datos y la administración de datos y objetos de SQL Server.
- ✓ **Integration Services** puede extraer y transformar datos de diversos orígenes como archivos de datos XML, archivos planos y orígenes de datos relacionales y, después, cargar los datos en uno o varios destinos.



SQL SERVER INTEGRATION SERVICE

Pregrado

Ingeniería de
Sistemas

- ✓ Integration Services incluye un amplio conjunto de tareas y transformaciones integradas, herramientas gráficas para crear paquetes y la base de datos del catálogo Integration Services, donde se almacenan, ejecutan y administran los paquetes.
- ✓ Puede usar las herramientas gráficas Integration Services para crear soluciones sin escribir ni una sola línea de código. También puede programar el amplio modelo de objeto Integration Services para crear paquetes mediante programación y codificar tareas personalizadas y otros objetos de paquete.



SQL SERVER INTEGRATION SERVICE

Pregrado

Ingeniería de
Sistemas

SQL Server Integration Service (SSIS) es un componente que permite generar procesos de migración de grandes cantidades de datos de diferentes orígenes llamados ETL.

Dispone de un entorno de desarrollo gráfico integrado en Visual Studio (para ello es necesario instalar los Data Tools). **SSIS** dispone de procesos que realizan “cosas” como ejecutar un Script SQL, leer datos de un fichero, leer datos de una tabla... y cada proceso se uno con otro mediante flujos de trabajo pudiendo comunicar ambos procesos.



Los proyectos **ETL** de **SSIS** tienen sentido cuando se pueden ejecutar y automatizar sin la necesidad de abrir el proyecto con Visual Studio, para ello, SSIS puede generar un paquete que se podrá ejecutar desde el agente de programación SQL Server Agent o crear un paquete ejecutable que y ejecutarlo desde línea de comandos.

SSIS es la evolución de los antiguos paquetes DTS y muchos DBA llaman a los nuevos paquetes de SSIS paquetes DTSX ya que esta es su extensión.



SQL SERVER INTEGRATION SERVICE

Pregrado

Ingeniería de
Sistemas

Uno de los inconvenientes de **SSIS** es que los paquetes desarrollados con una versión actual no se pueden ejecutar ni editar en versiones anteriores y paquetes diseñados con versiones anteriores, por lo general, se pueden ejecutar en versiones de SQL Server posteriores pero si el paquete falla en la ejecución, este debe editarse desde la versión correcta.



CARACTERISTICAS: SQL SERVER INTEGRATION SERVICE

- ✓ El **SSIS Import/Export Wizard** permite mover datos de origen a destino sin modificar los datos del origen y permitiendo hacer iteraciones y cambios de información antes de llegar al destino dentro de tablas de **ETL**. Se pueden importar datos de fuentes diferentes a SQL Server.
- ✓ Con la herramienta **Business Intelligence Development Studio**, se pueden realizar tareas de migración fácilmente usando tareas visuales. Si se desea crear nueva funcionalidad, se pueden crear scripts en C# o Visual Basic.
- ✓ Los paquetes, que son las unidades de almacenamiento de estas tareas de migración se pueden guardar en archivos dtsx o en la base de datos en formato **XML**.
- ✓ Una vez implementado el paquete puede ser depurado.



CARACTERISTICAS: SQL SERVER INTEGRATION SERVICE

- ✓ El **SSIS Import/Export Wizard** permite mover datos de origen a destino sin modificar los datos del origen y permitiendo hacer iteraciones y cambios de información antes de llegar al destino dentro de tablas de **ETL**. Se pueden importar datos de fuentes diferentes a SQL Server.
- ✓ Con la herramienta **Business Intelligence Development Studio**, se pueden realizar tareas de migración fácilmente usando tareas visuales. Si se desea crear nueva funcionalidad, se pueden crear scripts en C# o Visual Basic.
- ✓ Los paquetes, que son las unidades de almacenamiento de estas tareas de migración se pueden guardar en archivos dtsx o en la base de datos en formato **XML**.
- ✓ Una vez implementado el paquete puede ser depurado.



PROYECTO DE INTEGRATION SERVICE

Crear un proyecto

Plantillas de proyecto recientes

- | | | |
|--|--|--------------|
| | Biblioteca de clases (.NET Framework) | C# |
| | Aplicación de Windows Forms (.NET Framework) | C# |
| | Biblioteca de clases | C# |
| | Aplicación de Windows Forms | C# |
| | Integration Services Project | |
| | Solución en blanco | |
| | Aplicación de Windows Forms (.NET Framework) | Visual Basic |

integra

[Borrar todo](#)

C#

Windows

Biblioteca

No se ha encontrado ninguna coincidencia exacta.

Otros resultados basados en su búsqueda



Integration Services Project

This project may be used for building high performance data **integration** and workflow solutions that can be run on SSIS catalog, including extraction, transformation, and loading (ETL) operations for data warehousing.



Integration Services Project (Azure-Enabled)

This project may be used for building high performance data **integration** and workflow solutions that can also be run/ debugged on SSIS Platform-as-a-Service (PaaS) in Azure Data Factory.



Integration Services Import Project Wizard

A wizard that assists you in creating a new **Integration Services** (SSIS) project that is based on an existing one. Import from a project deployment file (.ispac extension) or from an **Integration Services** catalog on an instance of SQL Server.



PROYECTO DE INTEGRATION SERVICE

Configure su nuevo proyecto


Integration Services Project

Nombre del proyecto

Proyecto ETL SQL Server

Ubicación

C:\Users\IVAN\source\repos

Nombre de la solución 

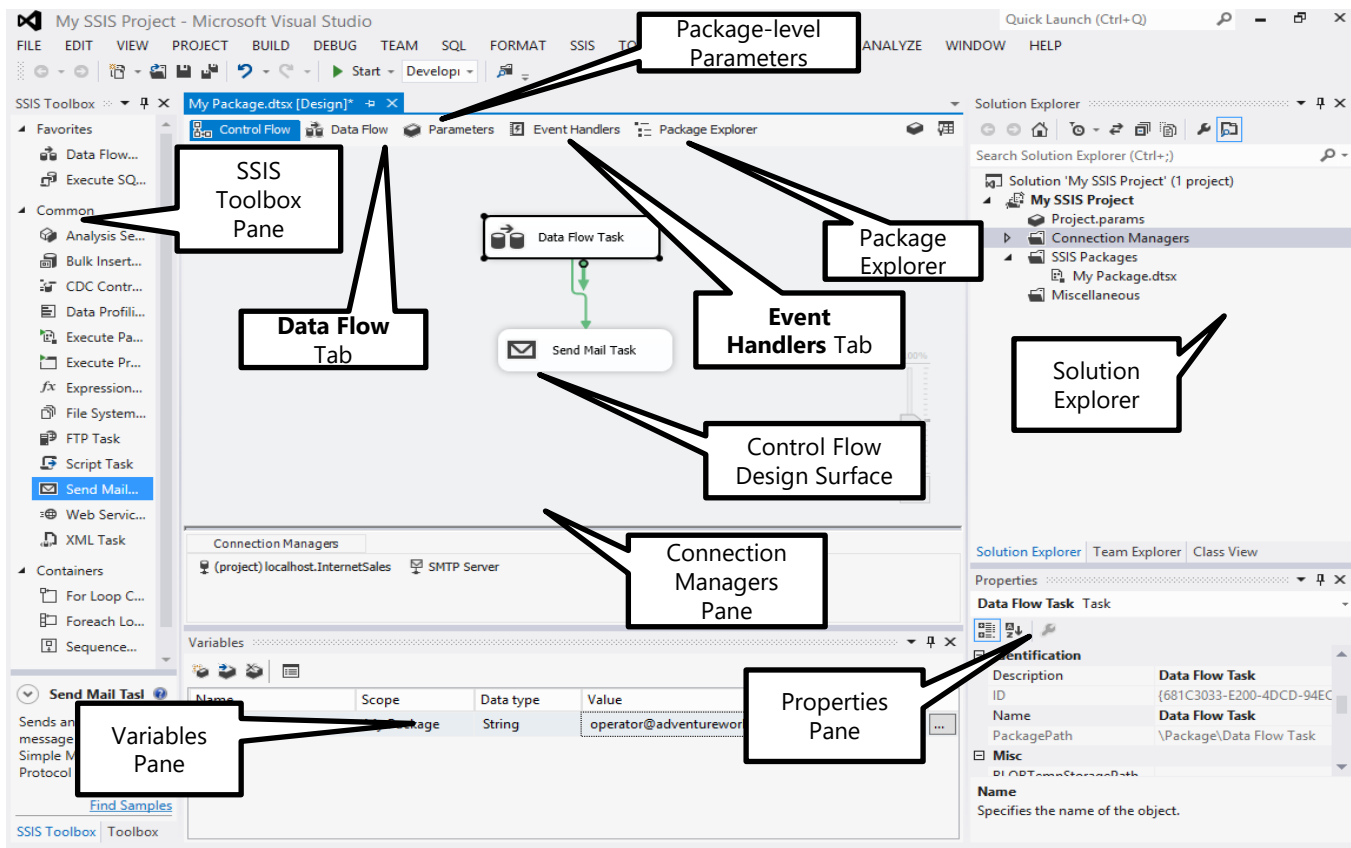
Proyecto ETL SQL Server



Colocar la solución y el proyecto en el mismo directorio



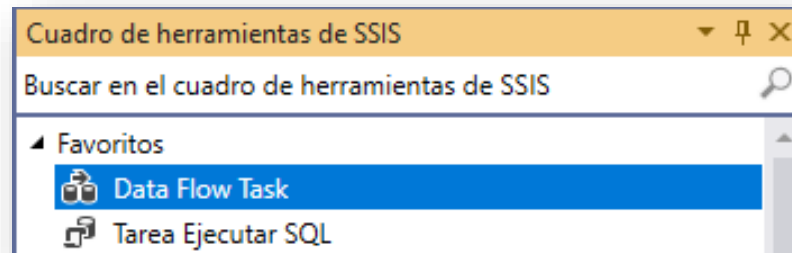
PROYECTO DE INTEGRATION SERVICE





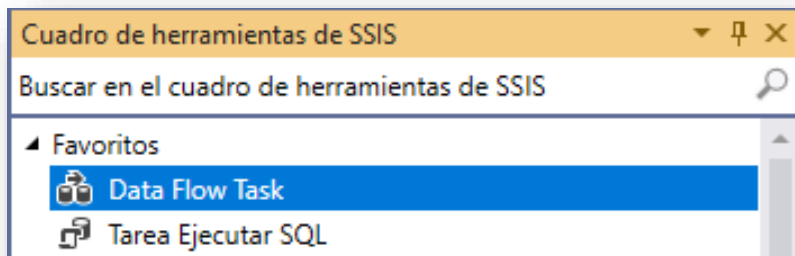
SSIS: TAREA FLUJO DE DATOS (DATA FLOW TASK)

- ✓ Encapsula el motor de flujo de datos que mueve datos entre orígenes y destinos, y permite al usuario transformar, limpiar y modificar datos a medida que se mueven. Agregar una tarea Flujo de datos a un flujo de control de paquetes permite que el paquete extraiga, transforme y cargue datos.
- ✓ Un flujo de datos se compone de por lo menos un componente de flujo de datos, pero normalmente es un conjunto de componentes de flujo de datos conectados: orígenes que extraen datos, transformaciones que modifican, enrutan o resumen datos, y destinos que cargan datos.





SSIS: TAREA FLUJO DE DATOS (DATA FLOW TASK)



- ✓ En el tiempo de ejecución, la tarea Flujo de datos genera un plan de ejecución a partir del flujo de datos y el motor de flujo de datos ejecuta el plan. Se puede crear una tarea Flujo de datos que no tenga flujo de datos, pero la tarea solo se ejecuta si incluye por lo menos un flujo de datos.
- ✓ Para realizar una inserción masiva de datos de archivos de texto en una base de datos de SQL Server , puede usar la tarea Inserción masiva en lugar de la tarea Flujo de datos y un flujo de datos. Sin embargo, la tarea Inserción masiva no transforma datos.



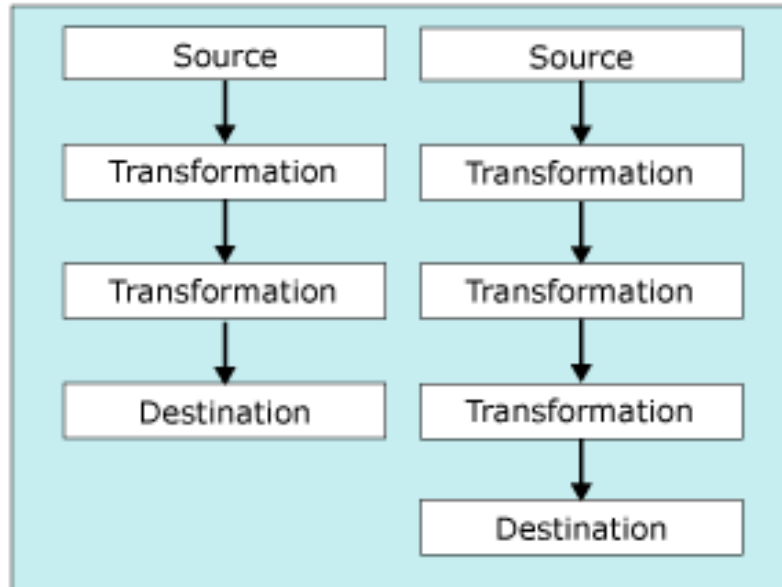
SSIS: TAREA FLUJO DE DATOS (DATA FLOW TASK)

- ✓ Una **tarea Flujo de datos** puede incluir varios flujos de datos. Si una tarea copia varios conjuntos de datos, y si el orden en que los datos se copian no es significativo, puede ser más conveniente incluir varios flujos de datos en la tarea Flujo de datos. **Por ejemplo**, puede crear cinco flujos de datos, y cada uno de ellos copiaría datos de un archivo plano en una tabla de dimensiones diferente en un esquema de estrella de almacenamiento de datos.
- ✓ Sin embargo, el motor de flujo de datos determina el orden de ejecución cuando existen varios flujos de datos dentro de una tarea de flujo de datos. Por tanto, cuando el orden es importante, el paquete debe usar varias tareas Flujo de datos, cada una de las cuales contiene un flujo de datos. En ese caso, puede aplicar restricciones de precedencia para controlar el orden de ejecución de las tareas.



SSIS: TAREA FLUJO DE DATOS (DATA FLOW TASK)

El siguiente diagrama muestra una tarea Flujo de datos con varios flujos de datos.





SSIS: FLUJO DE DATOS (DATAFLOW)

Pregrado

Ingeniería de
Sistemas

SQL Server Integration Services proporciona tres tipos diferentes de componentes de flujo de datos: orígenes, transformaciones y destinos. Los orígenes extraen datos de almacenes de datos tales como tablas y vistas en bases de datos relacionales, archivos y bases de datos de Analysis Services . Las transformaciones modifican, resumen y limpian datos. Los destinos cargan datos en almacenes de datos o crean conjuntos de datos almacenados en la memoria.

Además, **Integration Services** proporciona rutas que conectan la salida de un componente con la entrada de otro componente. Las rutas definen la secuencia de los componentes y permiten agregar anotaciones en el flujo de datos o ver el origen de la columna.



SSIS: FLUJO DE DATOS (DATAFLOW)

Los componentes de flujo de datos se conectan conectando la salida de orígenes y destinos con la entrada de transformaciones y destinos.

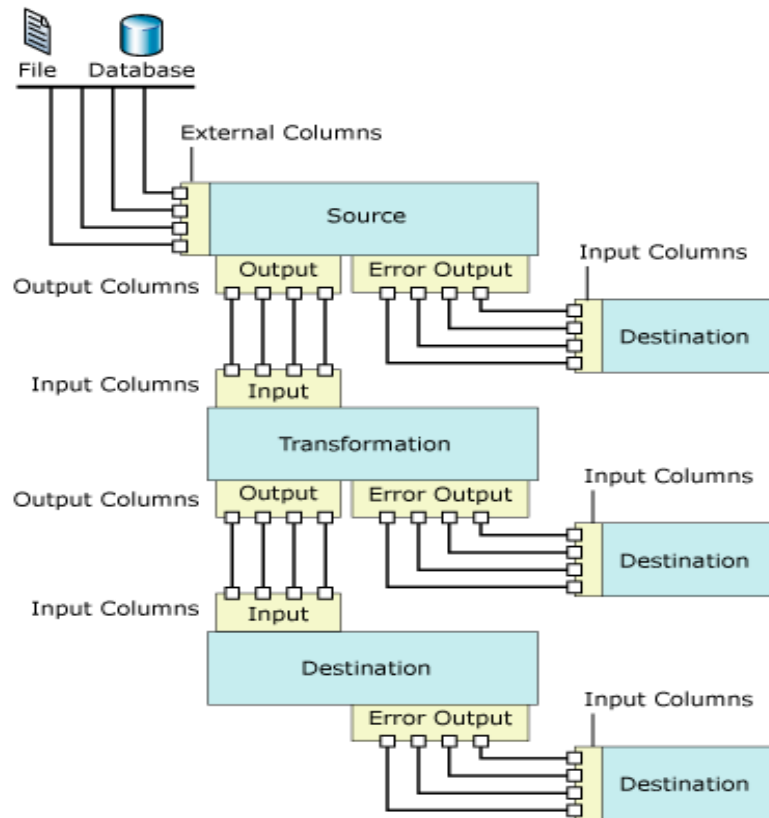
Al generar un flujo de datos normalmente se conecta el segundo componente y los componentes subsiguientes a medida que se agregan al flujo de datos. Después de conectar el componente, las columnas de entrada están disponibles para su uso en la configuración del componente.

Cuando no hay columnas de entrada disponibles, tiene que completar la configuración del componente después de conectarse al flujo de datos.



SSIS: FLUJO DE DATOS (DATAFLOW)

El siguiente diagrama muestra un flujo de datos que tiene un origen, una transformación con una entrada y una salida, y un destino. El diagrama incluye las entradas, salidas y salidas de error además de las columnas de entrada, salida y externas.





SSIS: IMPLEMENTACIÓN FLUJO DE DATOS (DATAFLOW)

- ✓ Agregar una tarea Flujo de datos al flujo de control de un paquete es el primer paso en la implementación de un flujo de datos en un paquete.
- ✓ Un paquete puede incluir varias tareas Flujo de datos, cada una de las cuales tiene su propio flujo de datos. **Por ejemplo**, si un paquete requiere que los flujos de datos se ejecuten en un orden especificado, o que se realicen otras tareas entre los flujos de datos, debe utilizar una tarea Flujo de datos independiente para cada flujo de datos.
- ✓ Una vez el flujo de control incluye una tarea Flujo de datos, puede empezar a generar el flujo de datos que usa un paquete.



SSIS: IMPLEMENTACIÓN FLUJO DE DATOS (DATAFLOW)

Crear un flujo de datos incluye las siguientes tareas:

- ✓ Agregar uno o más orígenes para extraer datos de archivos y bases de datos, y agregar administradores de conexión para conectarse a los orígenes.
- ✓ Agregar las transformaciones que satisfacen los requisitos empresariales del paquete. No es obligatorio que un flujo de datos incluya transformaciones.
- ✓ Algunas transformaciones requieren un administrador de conexiones. Por ejemplo, la transformación Búsqueda usa un administrador de conexión para conectarse a la base de datos que contiene los datos de búsqueda.
- ✓ Conectar componentes de flujo de datos conectando la salida de orígenes y transformaciones con la entrada de transformaciones y destinos.



SSIS: IMPLEMENTACIÓN FLUJO DE DATOS (DATAFLOW)

Crear un flujo de datos incluye las siguientes tareas:

- ✓ Agregar uno o varios destinos para cargar datos en almacenes de datos tales como archivos y bases de datos, y agregar administradores de conexiones para conectarse a los orígenes de datos.
- ✓ Configurar salidas de error en componentes para controlar problemas.
- ✓ En el tiempo de ejecución, pueden generarse errores de nivel de fila cuando los componentes de flujo de datos convierten datos, realizan una búsqueda o evalúan expresiones. **Por ejemplo**, una columna de datos con un valor de cadena no puede convertirse en un número entero, o una expresión intenta dividirse por cero. Ambas operaciones generan errores, y las filas que contienen los errores se pueden procesar independientemente mediante un flujo de errores.
- ✓ Incluir anotaciones para que el flujo de datos se autodocumente.



Las capacidades de las transformaciones presentan amplias variaciones. **Las transformaciones** pueden realizar tareas tales como ***actualizar, resumir, limpiar, combinar y distribuir datos***. Puede modificar valores en columnas, buscar valores en tablas, limpiar datos y agregar valores de columna.

Las entradas y salidas de una transformación definen las columnas de datos de entrada y salida. Según la operación realizada con los datos, algunas transformaciones tienen una sola entrada y varias salidas, mientras que otras transformaciones tienen varias entradas y una sola salida.

Las transformaciones también pueden incluir salidas de error que proporcionan información sobre el error que se ha producido, junto con los datos en los que se ha producido un error; por ejemplo, los datos de cadena que no se han podido convertir a un tipo de datos enteros.



El modelo de objetos de **Integration Services** no restringe la cantidad de entradas, salidas normales y salidas de error que pueden contener las transformaciones. Puede crear transformaciones personalizadas que implementan cualquier combinación de varias entradas, salidas normales y salidas de error.

La entrada de una transformación se define como una o más columnas de entrada. Algunas transformaciones de **Integration Services** también pueden hacer referencia a columnas externas como entrada.

Por ejemplo, la entrada de la transformación Comando de OLE DB incluye columnas externas. Una columna de salida es una columna que la transformación agrega al flujo de datos. Tanto las salidas normales como las salidas de error contienen columnas de salida. Estas columnas de salida a su vez funcionan como columnas de entrada para el siguiente componente en el flujo de datos, ya sea otra transformación o un destino.



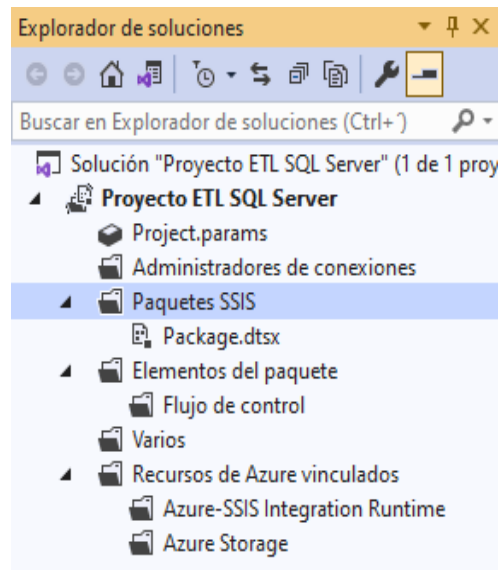
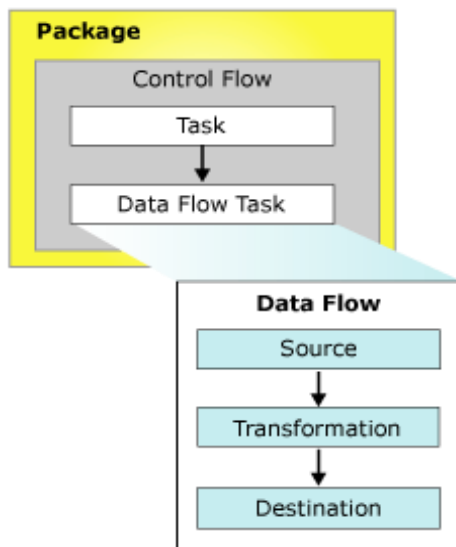
Las siguientes transformaciones tienen propiedades que se pueden actualizar a través de expresiones de propiedad:

- ✓ Transformación División condicional
- ✓ Transformación Columna derivada
- ✓ Agrupación aproximada, transformación
- ✓ Búsqueda aproximada, transformación
- ✓ Transformación Comando de OLE DB
- ✓ Transformación Muestreo de porcentaje
- ✓ Transformación dinámica
- ✓ Transformación Muestreo de fila
- ✓ Transformación Ordenar
- ✓ Transformación Anular dinamización



PAQUETE INTEGRATION SERVICE (SSIS)

Un paquete es una colección organizada de conexiones, elementos de flujo de control, elementos de flujo de datos, controladores de eventos, variables, parámetros y configuraciones que se pueden ensamblar con la ayuda de las herramientas de diseño gráfico proporcionadas por **SQL Server Integration Services** o mediante programación.





CONTENIDO DE PAQUETE INTEGRATION SERVICE (SSIS)

Tareas y contenedores (flujo de control). Un flujo de control consta de una o más tareas y contenedores que se ejecutan cuando se ejecuta el paquete. Para controlar el orden o definir las condiciones para ejecutar la siguiente tarea o contenedor del flujo de control del paquete, puede usar restricciones de precedencia para conectar las tareas y los contenedores de un paquete. También se puede agrupar y ejecutar repetidamente un subconjunto de tareas y contenedores como una unidad en el flujo de control del paquete





CONTENIDO DE UN PAQUETE INTEGRATION SERVICE (SSIS)

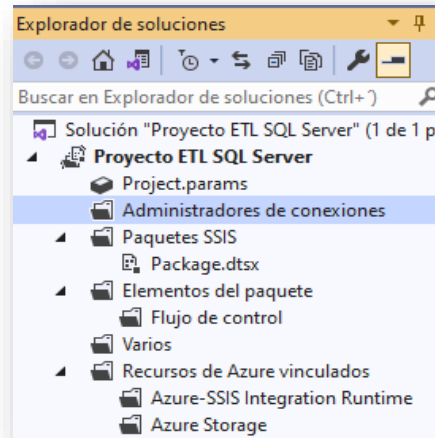
Orígenes de datos y destinos (flujo de datos). Un flujo de datos consta de los orígenes y destinos que extraen y cargan datos, las transformaciones que modifican y extienden datos, y las rutas que vinculan orígenes, transformaciones y destinos. Para poder agregar un flujo de datos a un paquete, el flujo de control de paquetes debe incluir una tarea Flujo de datos. La tarea Flujo de datos es el ejecutable del paquete SSIS que crea, organiza y ejecuta el flujo de datos. Se abre una instancia independiente del motor de flujo de datos para cada tarea Flujo de datos de un paquete.





CONTENIDO DE UN PAQUETE INTEGRATION SERVICE (SSIS)

Administradores de conexiones (conexiones). Un paquete suele incluir al menos un administrador de conexiones. Un administrador de conexiones es un vínculo entre un paquete y un origen de datos que define la cadena de conexión para acceder a los datos que las tareas, transformaciones y controladores de eventos del paquete usan. **Integration Services** incluye tipos de conexiones para orígenes de datos tales como archivos de texto y XML, bases de datos relacionales y proyectos y bases de datos de Analysis Services .



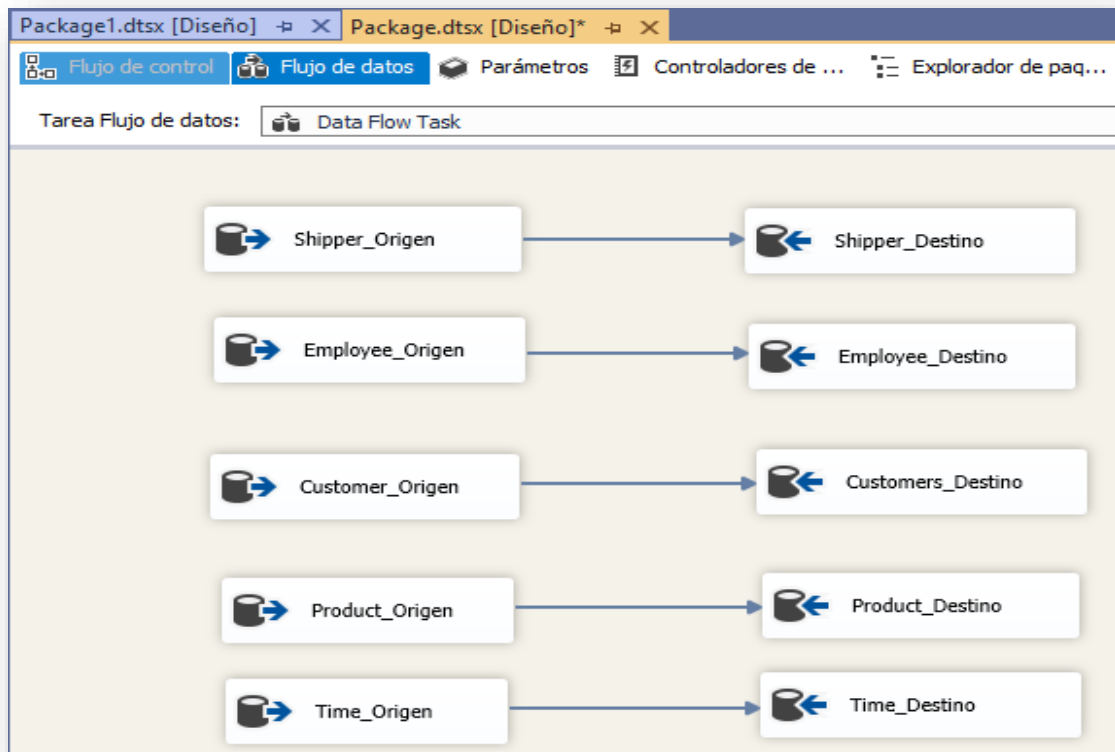


ETL (EXTRACCIÓN , TRANSFORMACIÓN Y CARGA)





ETL DATAINTEGRATION DATAMART NORTHWIND





ETL DATAINTEGRATION DATAMART NORTHWIND

Editor de origen de OLE DB

Configure las propiedades que utiliza un flujo de datos para obtener datos de cualquier proveedor OLE DB.

Administrador de cone
Columnas
Salida de error

Especifique un administrador de conexiones OLE DB, un origen de datos o una vista del origen de datos y seleccione el modo de acceso a los datos. Si utiliza el modo de acceso a comandos SQL, especifique el comando SQL escribiendo la consulta o mediante el generador de consultas.

Administrador de conexiones OLE DB:
DESKTOP-LA7744P.Northwind Nueva...

Modo de acceso a datos:
Tabla o vista

Nombre de la tabla o la vista:
[dbo].[Vista_Shipper_Dim]

Vista previa...

Aceptar Cancelar Ayuda

CONFIGURAR EL
ORIGEN PARA CADA
VISTA CREADA EN
NORTHWIND



ETL DATAINTEGRATION DATAMART NORTHWIND

Pregrado

Ingeniería de
Sistemas

Editor de destino de OLE DB

Configure las propiedades para insertar datos en una base de datos relacional mediante un proveedor OLE DB.

Administrador de conexiones OLE DB:
Asignaciones
Salida de error

Especifique un administrador de conexiones OLE DB, un origen de datos o una vista del origen de datos y seleccione el modo de acceso a los datos. Si utiliza el modo de acceso a comandos SQL, especifique el comando SQL escribiendo la consulta o mediante el generador de consultas. Para obtener acceso a datos de carga rápida, configure las opciones de actualización de tablas.

Administrador de conexiones OLE DB:
DESKTOP-LA7744P.NorthwindEstrella Nueva...

Modo de acceso a datos:
Carga rápida de tabla o vista

Nombre de la tabla o la vista:
[dbo].[Shipper_Dim] Nueva...

☐ Mantener valores de identidad ☒ Bloqueo de tabla
☒ Mantener valores NULL ☒ Comprobar restricciones

Filas por lote:

Tamaño máximo de confirmación de inserción: 2147483647

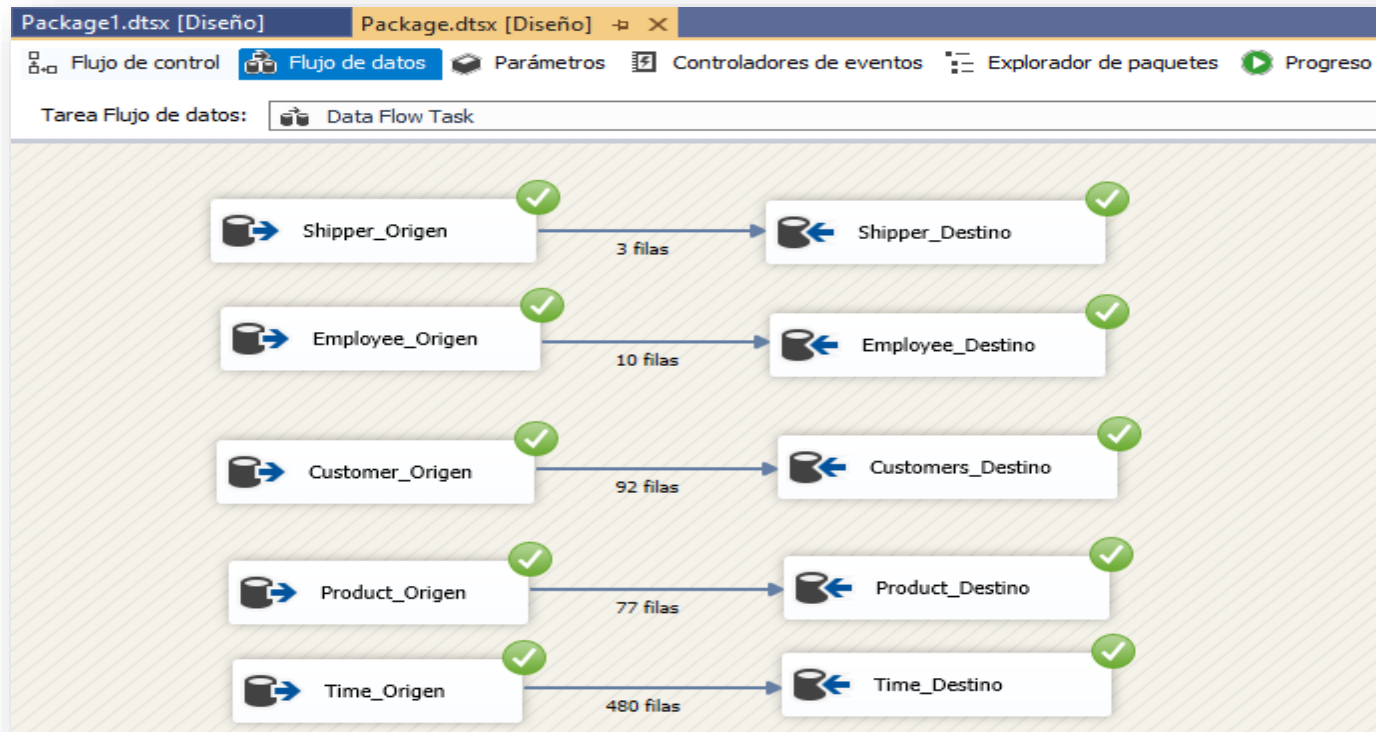
Ver datos

Aceptar Cancelar Ayuda

**CONFIGURAR EL
DESTINO PARA
CADA TABLA DE
DIMENSIONES
DEL DATAMART
DE NORTHWIND**



ETL DATAINTEGRATION DATAMART NORTHWIND



**EJECUTAR ETL
PARA LLENAR
LAS FILAS DE
LAS TABLAS
DE
DIMENSIONES**



ETL DATAINTEGRATION DATAMART NORTHWIND



**CONFIGURAR EL
ORIGEN Y EL DESTINO
PARA CADA TABLA
HECHOS DEL
DATAMART DE
NORTHWIND**

Con la ejecución del último paquete habremos terminado de llenar los registros de las tablas de dimensiones y de hechos del Datamart Northwind a través del proceso del ETL realizado en el DataIntegration.



¿QUÉ HEMOS APRENDIDO HOY?



Para que reflexionen y entiendan la importancia de los temas tratados y el mejoramiento de su propio proceso de aprendizaje.



Universidad **César Vallejo**

Licenciada por Sunedu
para que puedas salir adelante