

Universidad de Costa Rica

Facultad de Ingeniería

Escuela de Ciencias de la Computación e Informática
Bases de Datos Avanzadas

Actividad Asincrónica. Práctica Dirigida ETL

2025



Autor

Brandon Trigueros Lara C17899

Pasos previos: Conceptos a familiarizarse

Data Flow Task

Un Data Flow Task es un componente del paquete de Integration Service que me permite extraer datos de fuentes heterogéneas múltiples, realizar transformaciones sobre estos datos como derivación de columnas, limpieza, ordenamiento, etc, y finalmente cargar los datos transformados hacia distintos destinos [1]. Internamente realiza operaciones como verificar metadatos de los orígenes, la construcción de un grafo dirigido acíclico para el flujo de datos, entre otros. Todo esto dentro de un mismo paso lógico.

Execute SQL Task

Es una tarea de Control Flow que permite ejecutar sentencias SQL o procedimientos almacenados dentro de un paquete. Permite ejecutar comandos DML, DDL así como llamar a procedimientos almacenados [2].

File System Task

Es otra tarea de Control Flow, pero en lugar de gestionar memoria, esta vez dedicada a ejecución de operaciones sobre el gestor de archivos del sistema operativo sobre el que corre el paquete. Gestiona ficheros y archivos. Se usa para preparar directorios de staging antes de cargas, para archivar archivos ya procesados o ajustar atributos de los archivos de manera dinámica [3].

OLE DB Source

Es un componente de Data Flow que extrae datos de orígenes relacionales que soporten OLE DB. Funciona en el contexto de Data Flow Task, alimentando transformaciones y destinos. También permite usar el botón Preview para obtener hasta 200 filas de muestra y verificar que las columnas y filtros funcionan correctamente [4].

Derived Column

Es una transformación, permite calcular nuevas columnas o reemplazar valores existentes en función de otras variables, operaciones y columnas de entrada. Utiliza el lenguaje de expresiones de SSIS, el cual tiene

- Operadores aritméticos como +, -, *, / ,

- Funciones integradas como SUBSTRING(), CASE, ISNULL, etc

- Variables de paquete [5].

Aggregate

La transformación **Aggregate** de SSIS agrupa filas y aplica funciones agregadas (como Sum, Average, Count) sobre columnas de un flujo de datos, generando uno o más resultados que pueden direccionarse a distintos caminos de salida. Soporta las siguientes operaciones estándar: Group By, Sum, Average, Count y Count Distinct, Min y Max [6].

Sort

Sort me ordena ascendente o descendentemente los registros de un flujo de datos y los copia en un destino definido. Puede ordenar los datos en uno o más niveles y además permite eliminar filas duplicadas [7].

Data Viewer

Es una herramienta de depuración cuyo principal objetivo es la de mostrar, buffer por buffer ciertas columnas seleccionadas entre dos elementos de un Data Flow Task. Se muestran los datos en formato de gráficos o cuadrícula, lo cual permite identificar fácilmente valores atípicos para su pronto tratamiento [8].

Multiple Destinations

El componente Multicast en SSIS permite duplicar el flujo de datos entrante y enviarlo simultáneamente a múltiples destinos. Es útil cuando necesitas aplicar diferentes transformaciones o almacenar los mismos datos en distintos formatos o ubicaciones [9].

Flat File Destination

El componente Flat File Destination se utiliza para exportar datos a archivos de texto plano, como .txt o .csv. Es comúnmente usado para generar archivos que pueden ser fácilmente compartidos o procesados por otros sistemas [11].

Excel Destination

El componente Excel Destination permite exportar datos directamente a hojas de cálculo de Microsoft Excel. Es útil para generar informes o compartir datos en un formato ampliamente utilizado y fácil de leer [10].

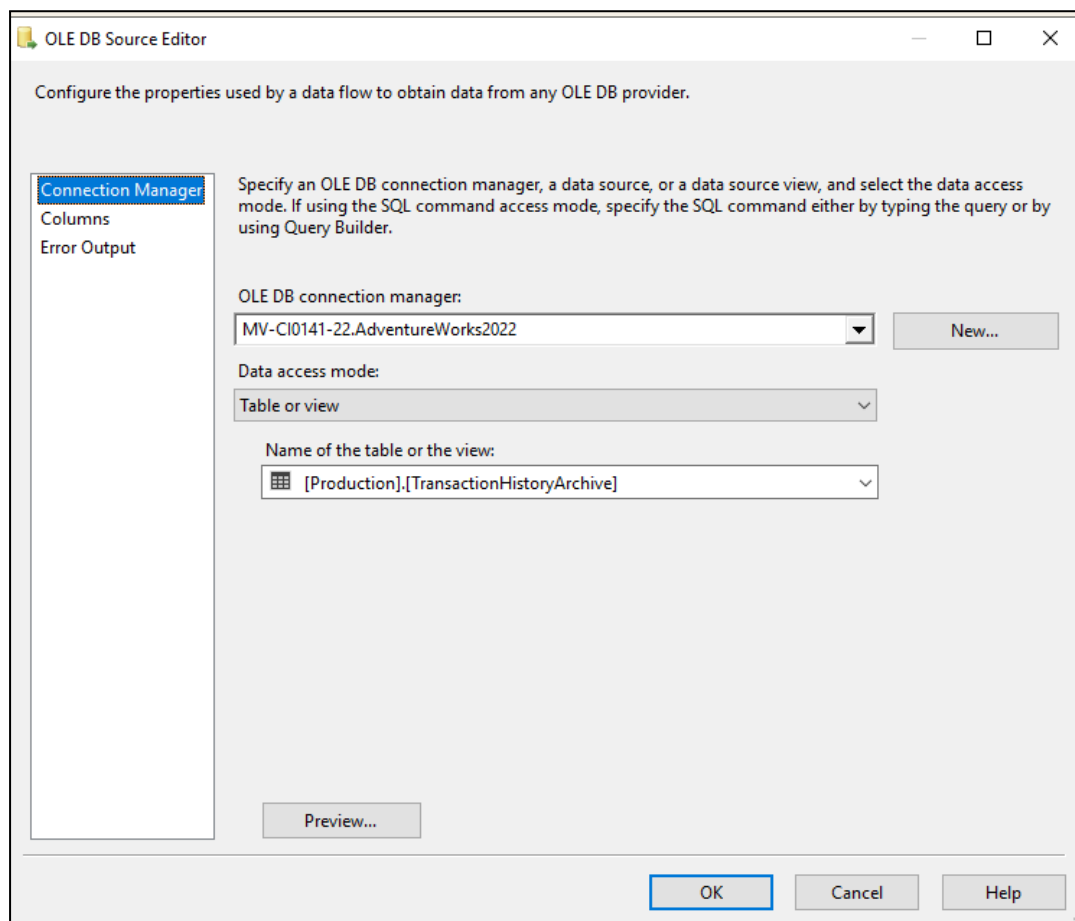
Abrir el puerto de la base de datos en el sistema operativo

Para permitir que otros equipos accedan a tu instancia de SQL Server desde la red, es necesario abrir el puerto correspondiente en el firewall del sistema operativo. Por defecto, SQL Server utiliza el puerto TCP 1433 para conexiones remotas. Los pasos para hacerlo son:

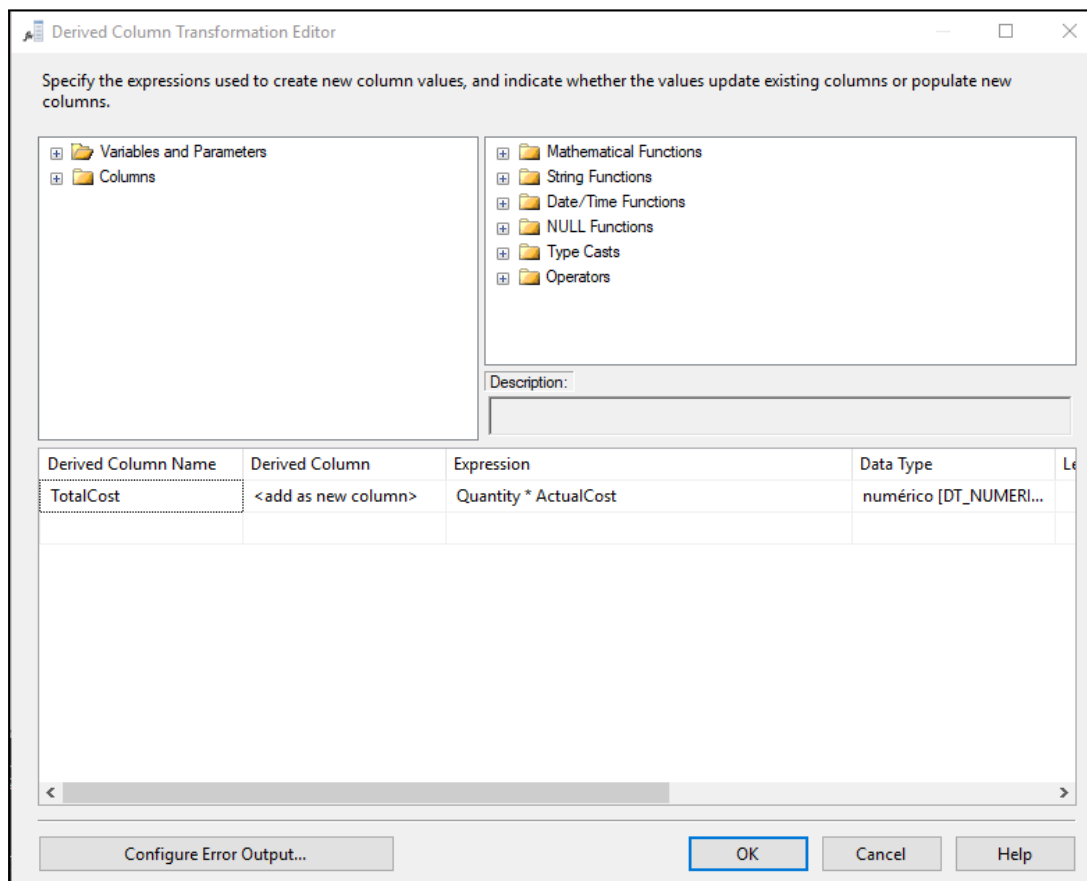
1. **Abrir el Firewall de Windows con seguridad avanzada:**
 - a. Presiona **Win + R**, escribe **wf.msc** y presiona Enter.
2. **Crear una nueva regla de entrada:**
 - a. En el panel izquierdo, selecciona **Reglas de entrada**.
 - b. En el panel derecho, haz clic en **Nueva regla...**
3. **Configurar la regla:**
 - a. Selecciona **Puerto** y haz clic en **Siguiente**.
 - b. Elige **TCP** y selecciona **Puertos locales específicos**.
 - c. Escribe **1433** y haz clic en **Siguiente**.
 - d. Selecciona **Permitir la conexión** y haz clic en **Siguiente**.
 - e. Marca los perfiles de red que correspondan (Dominio, Privado, Público) y haz clic en **Siguiente**.
 - f. Asigna un nombre a la regla, por ejemplo, "SQL Server TCP 1433", y haz clic en **Finalizar** [12].

Práctica ETL

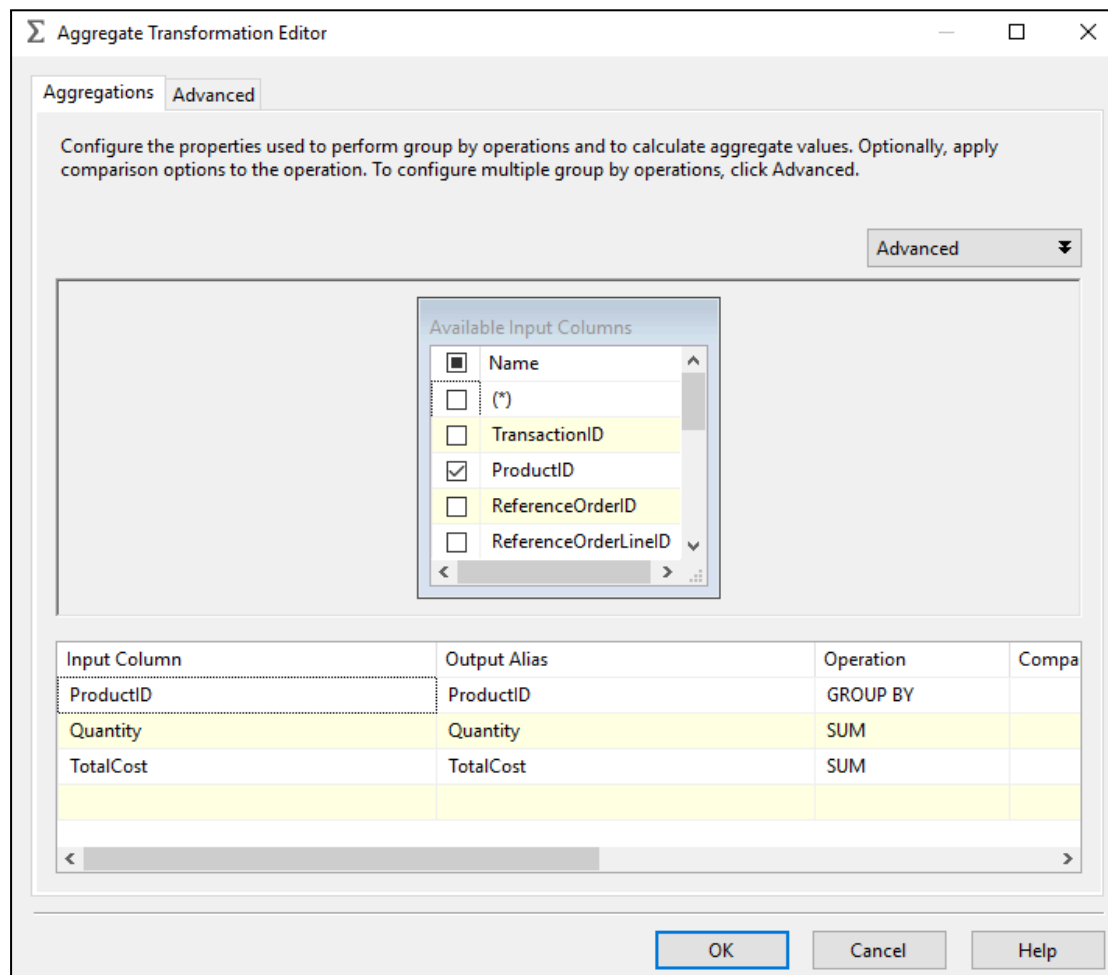
Configuramos la fuente de datos como un OLE DB Source para tomar la base de datos AdventureWorks2022:



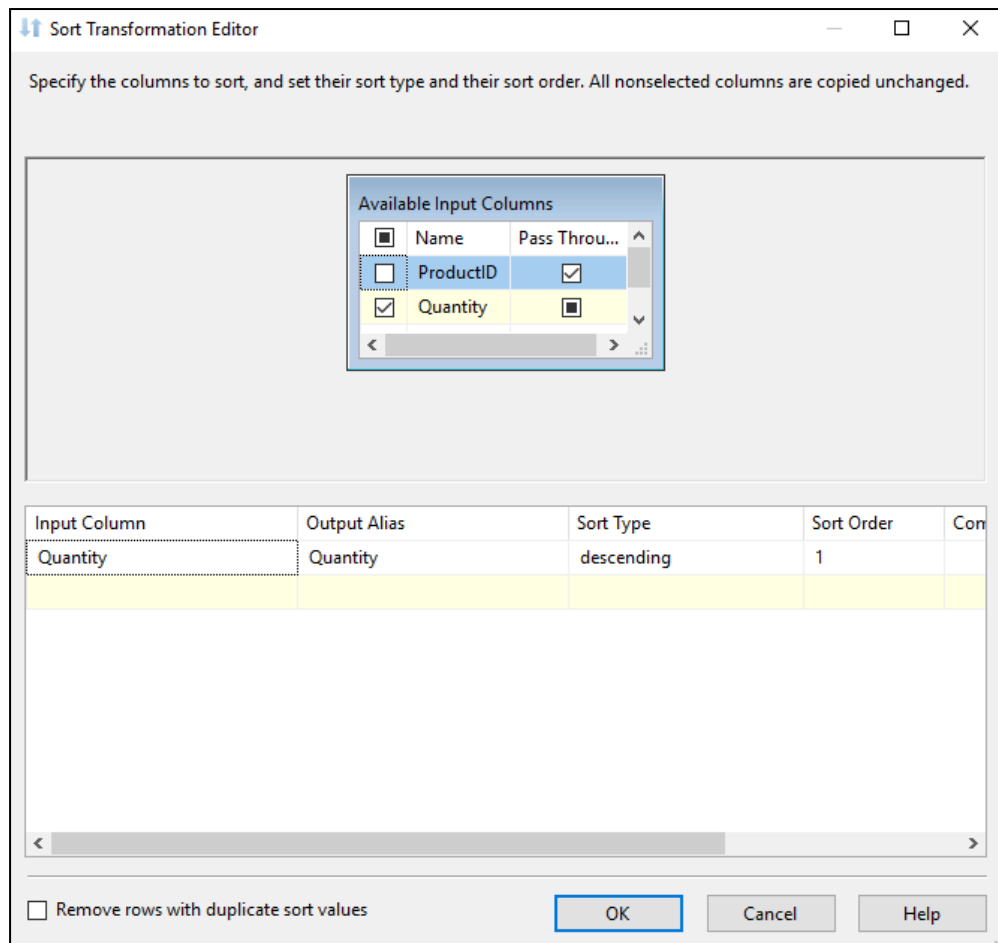
Aplicamos la transformación de columna derivada como se muestra a continuación



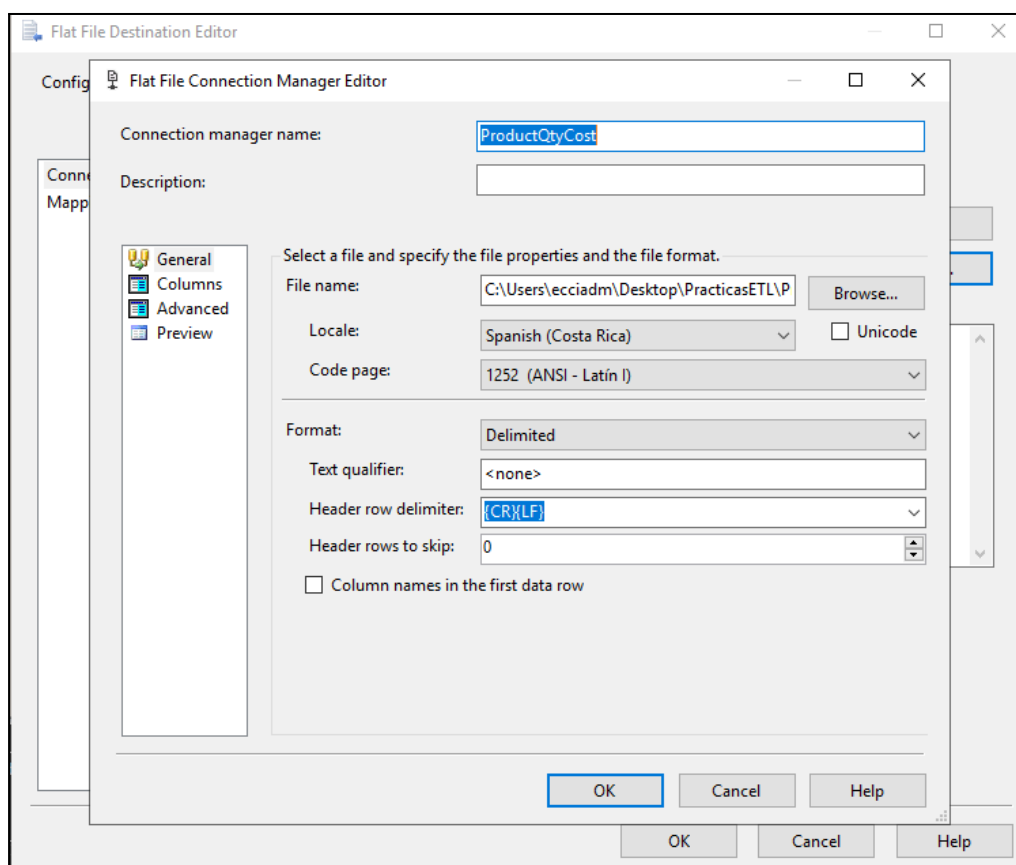
Seguidamente arrastramos y editamos la transformación de sumalización:



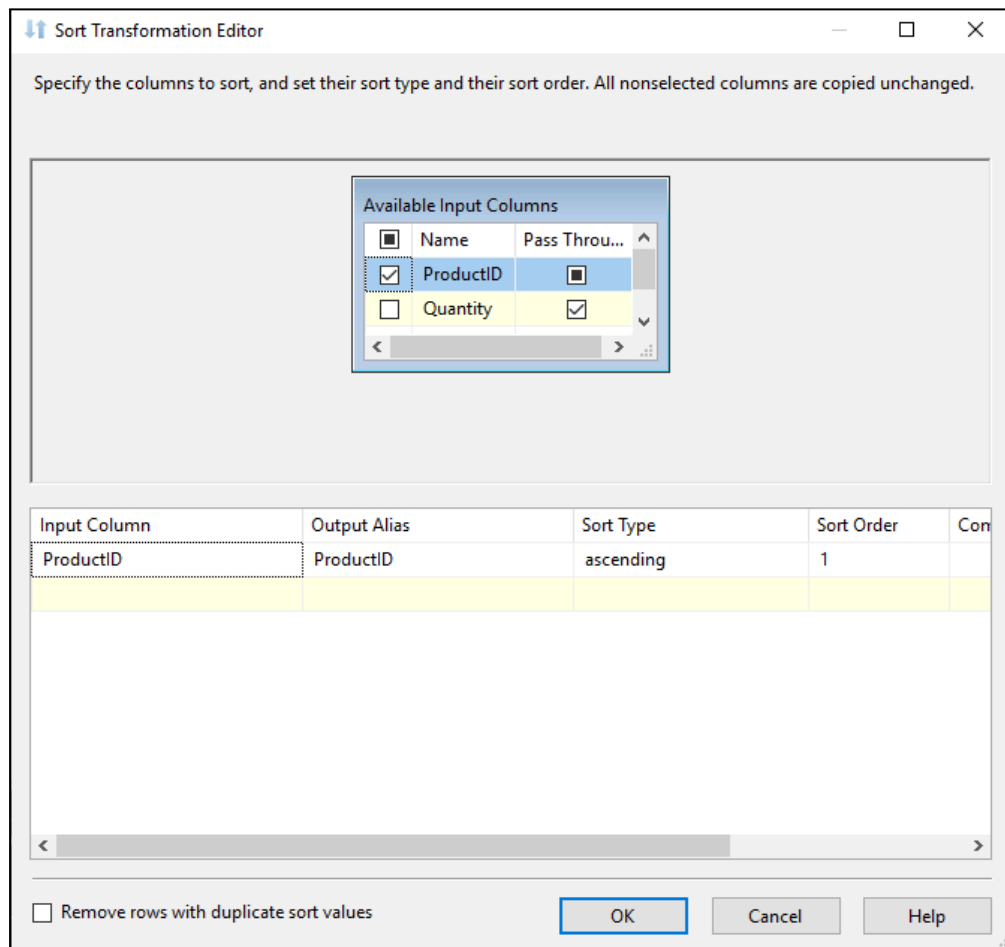
Usamos la transformación de Sort y la editamos para que ordene por la columna Cantidad:



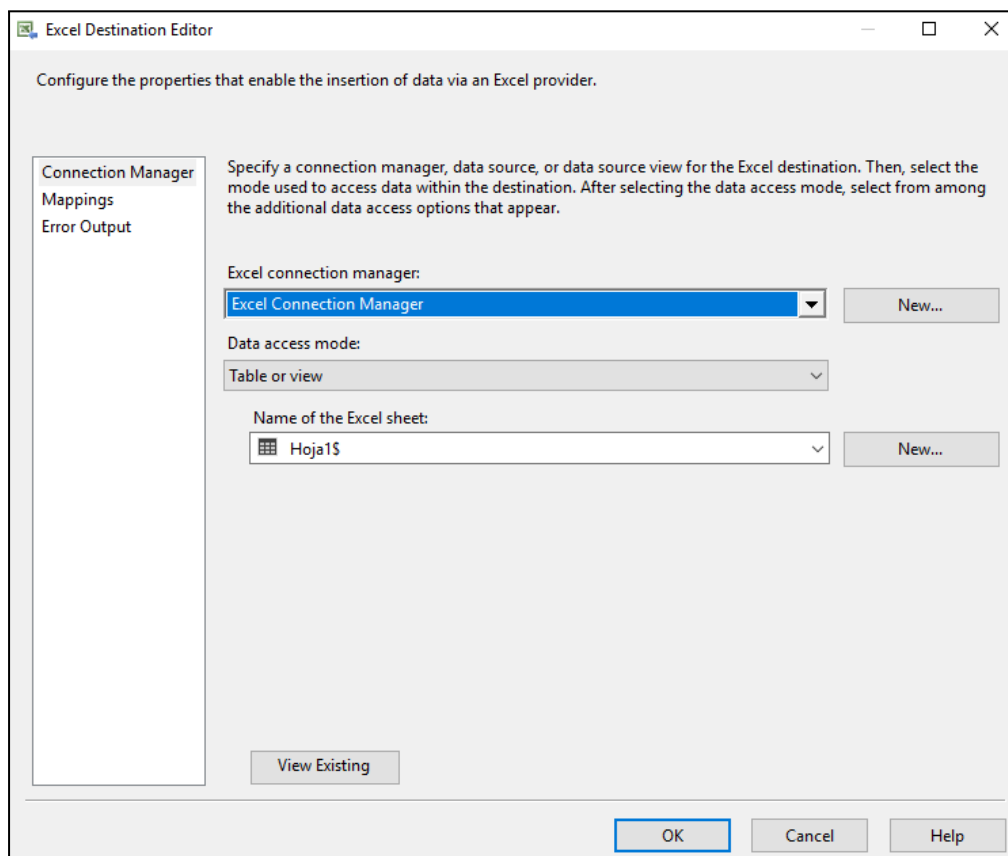
Configuración de destino a texto plano de la siguiente manera:



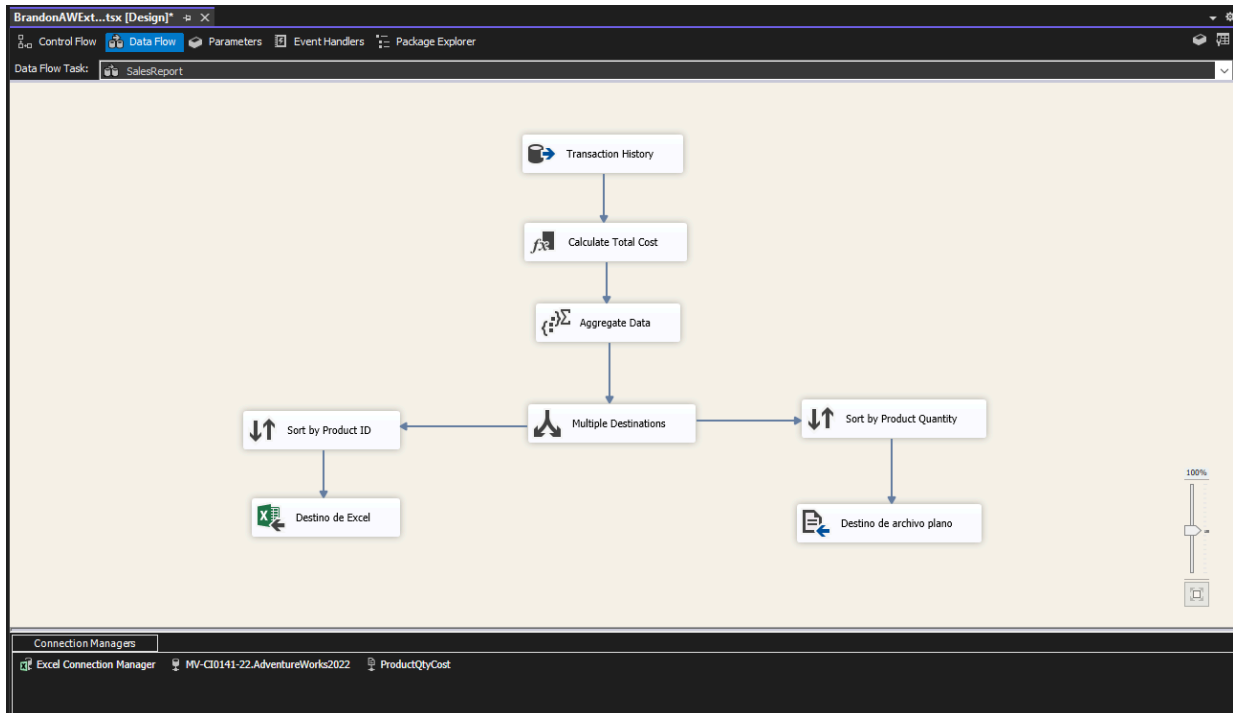
Similarmemente hacemos un Sort por ProductId:



Configuración de destino a Excel:

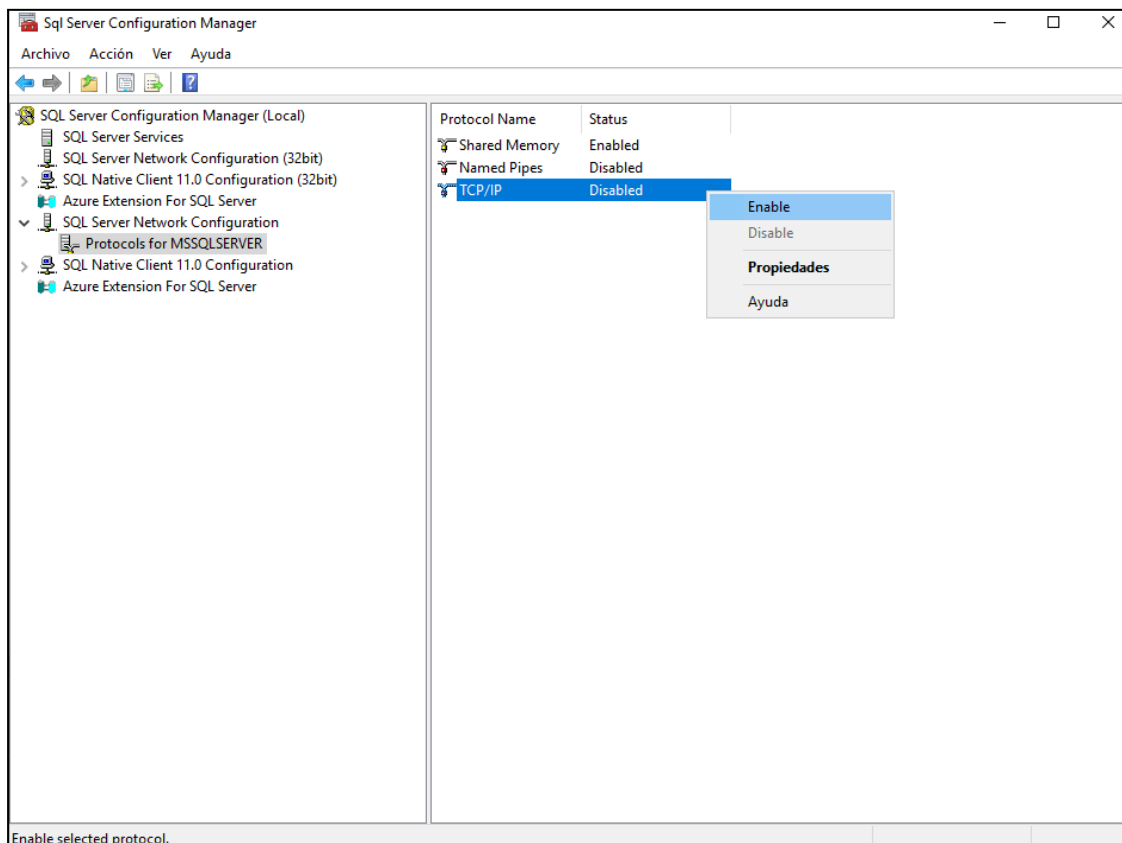


Esquema ETL completado, esta versión solamente se conectaba a la base de datos en la misma máquina:

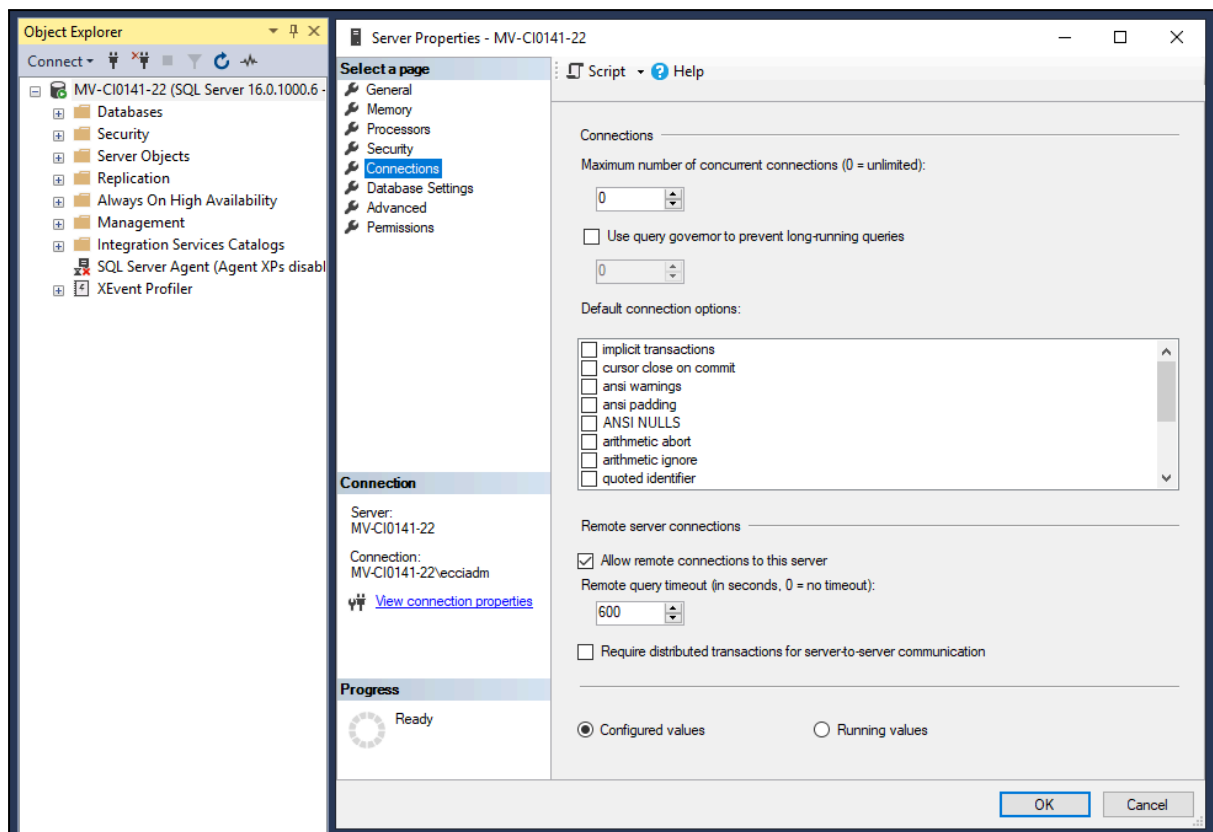


Conección de cliente ETL a servidor SQL

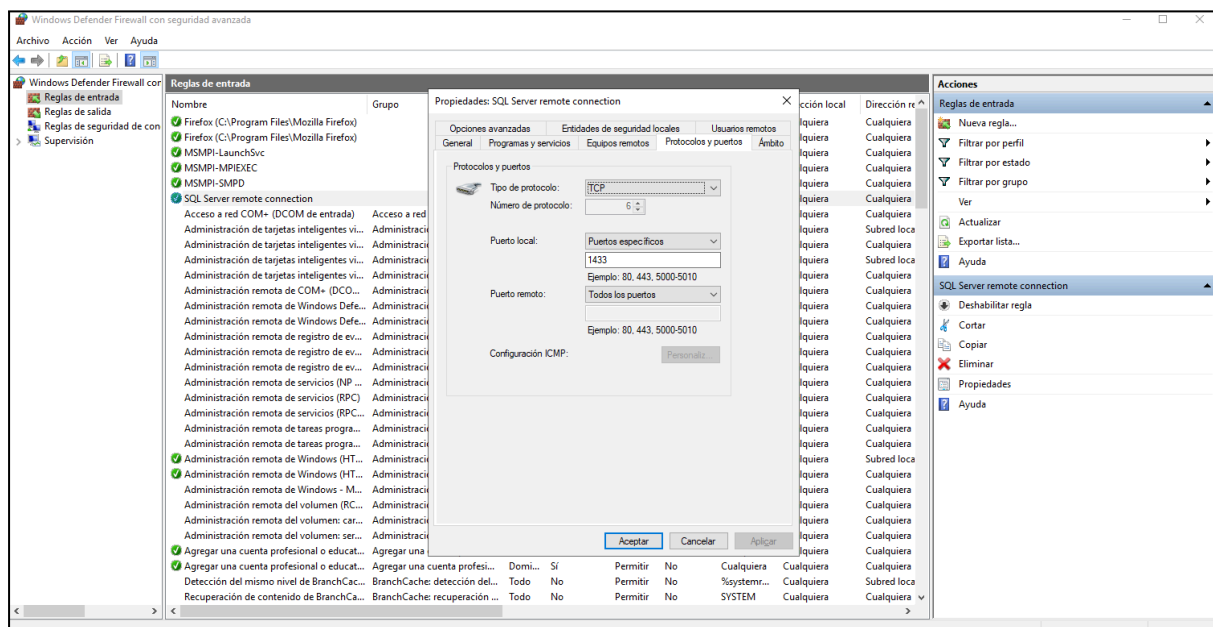
Server: Habilitar protocolo TCP/IP en Configuration Manager



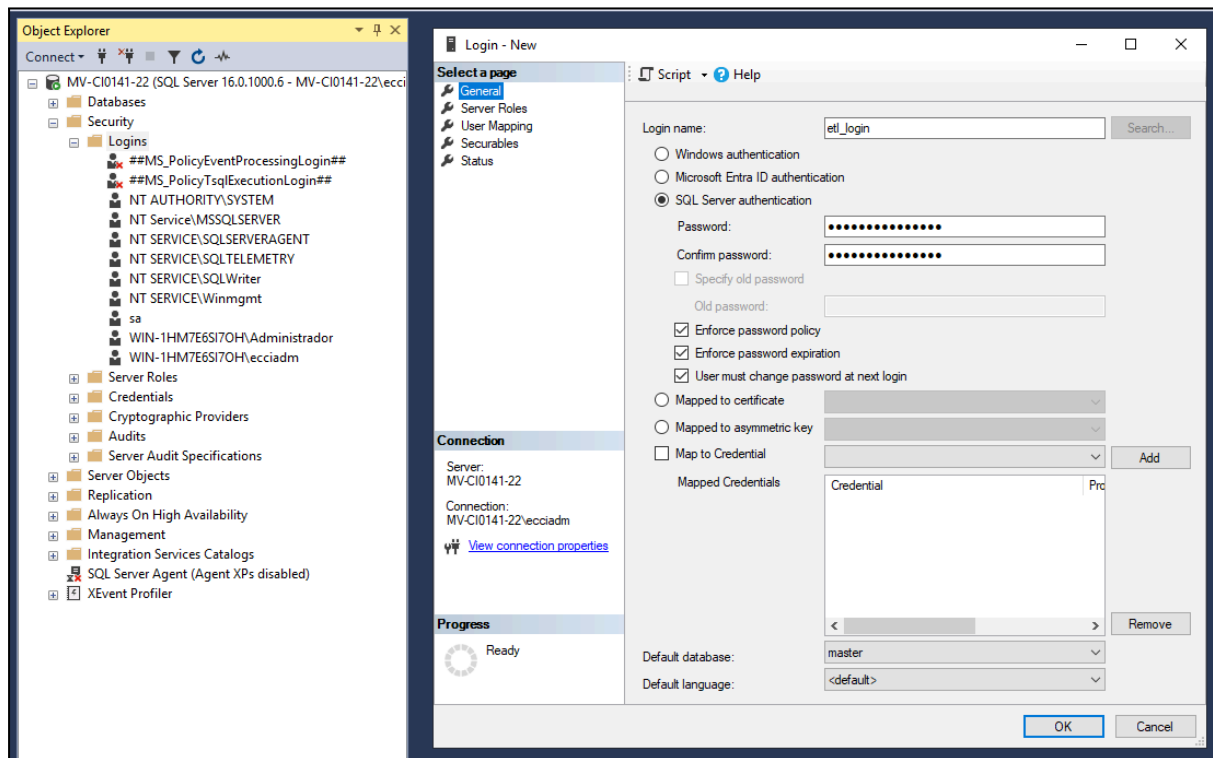
Server: Permitir conexiones remotas al servidor



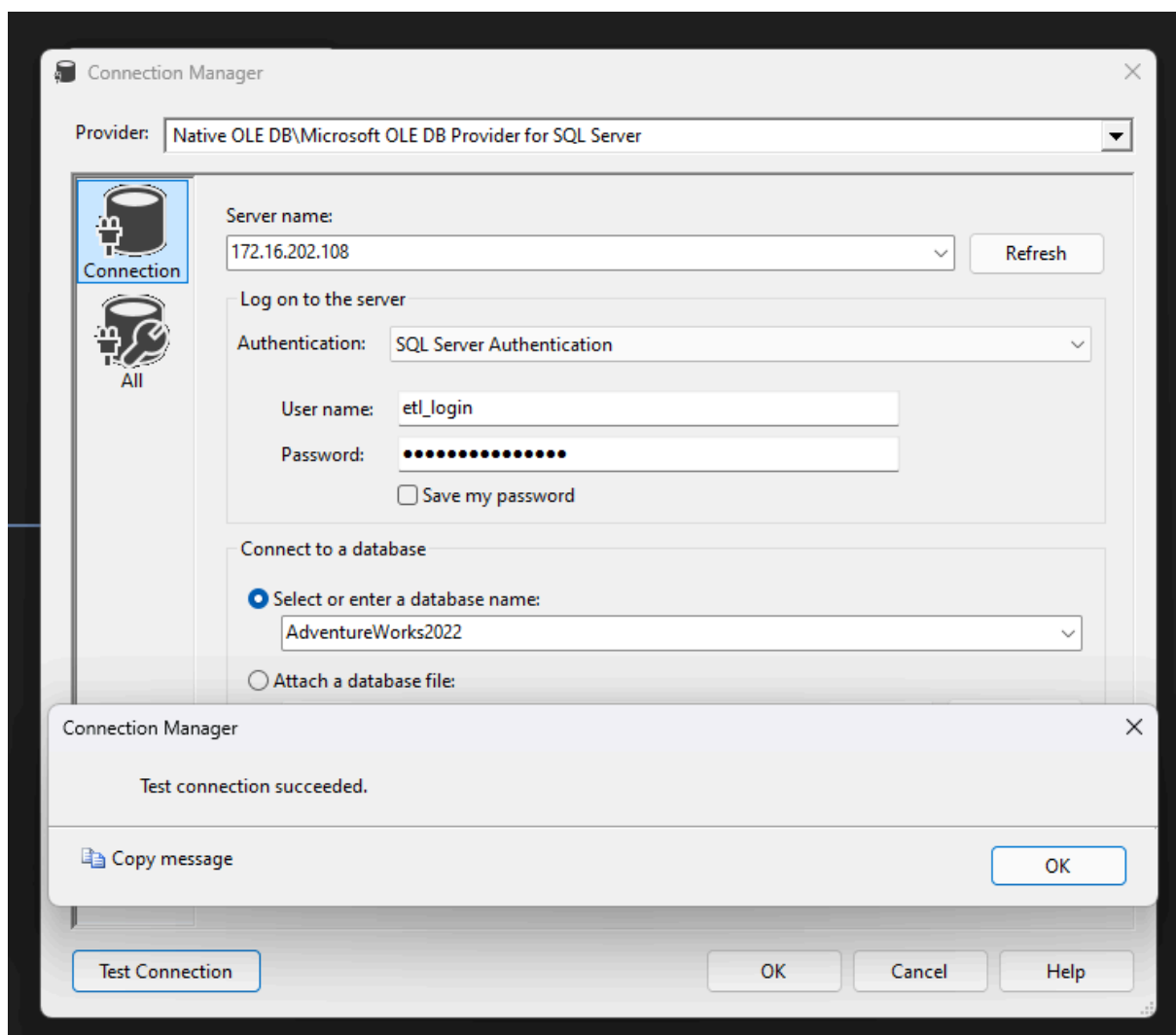
Servidor: Abrir puerto 1433



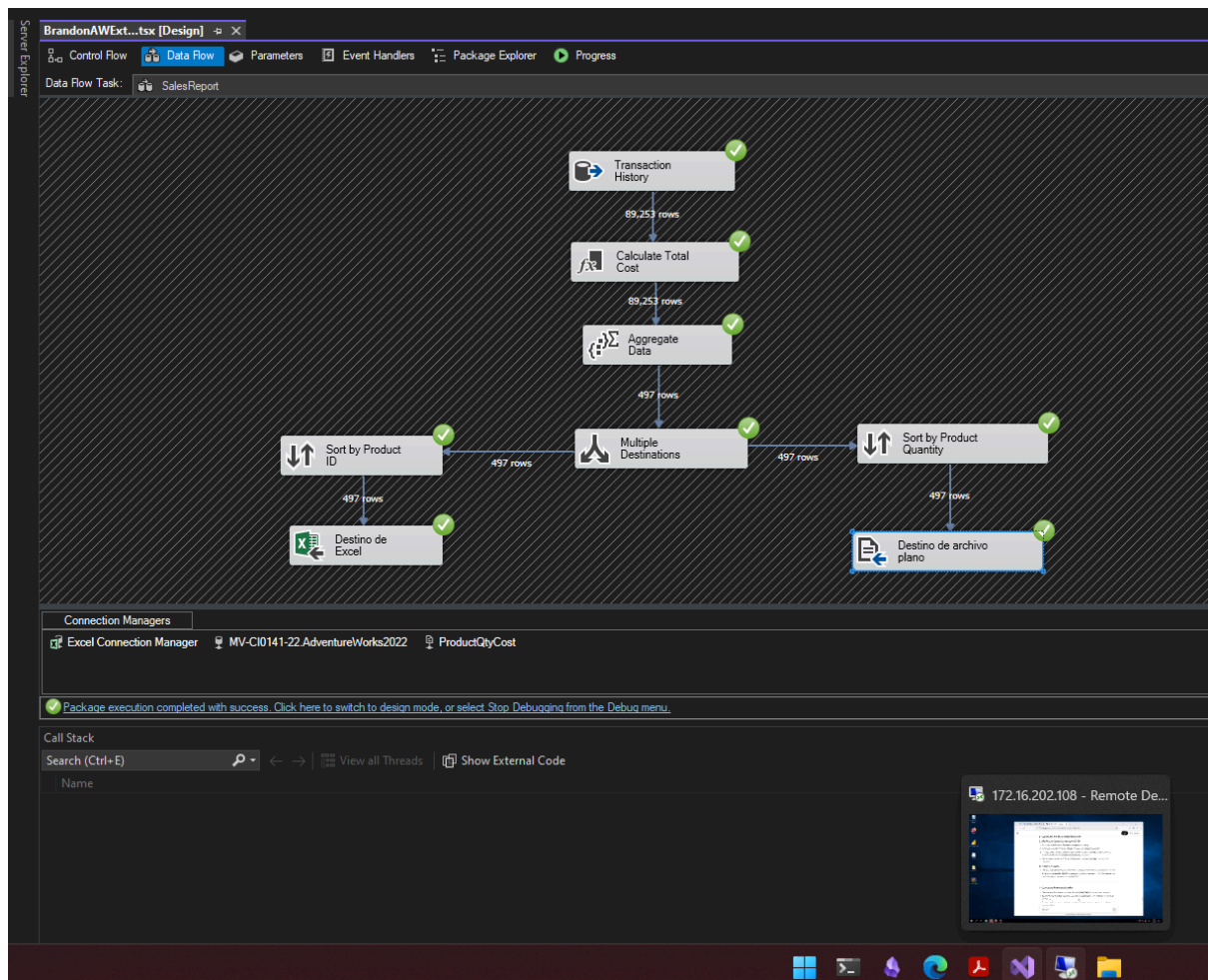
Server: Creación de usuario



Client: Editar conexión OLE DB



ETL corriendo desde cliente y conectando con servidor SQL



Referencias

Conversación ChatGPT

<https://chatgpt.com/share/680c1715-d32c-8010-814f-edb5c69fae7a>

[1] Microsoft, "Integration Services Tasks - SQL Server Integration Services (SSIS)," Microsoft Learn. [Online]. Available: <https://learn.microsoft.com/en-us/sql/integration-services/control-flow/integration-services-tasks>. [Accessed: 24-Apr-2025].

[2] Microsoft, "Execute SQL Task - SQL Server Integration Services (SSIS)," Microsoft Learn. [Online]. Available: <https://learn.microsoft.com/en-us/sql/integration-services/control-flow/execute-sql-task>. [Accessed: 24-Apr-2025].

[3] Microsoft, "File System Task - SQL Server Integration Services (SSIS)," Microsoft Learn. [Online]. Available: <https://learn.microsoft.com/en-us/sql/integration-services/control-flow/file-system-task>. [Accessed: 24-Apr-2025].

[4] Microsoft, "OLE DB Source - SQL Server Integration Services (SSIS)," Microsoft Learn. [Online]. Available: <https://learn.microsoft.com/en-us/sql/integration-services/data-flow/ole-db-source>. [Accessed: 24-Apr-2025].

[5] Microsoft, "Derived Column Transformation - SQL Server Integration Services (SSIS)," Microsoft Learn. [Online]. Available: <https://learn.microsoft.com/en-us/sql/integration-services/data-flow/derived-column-transformation>. [Accessed: 24-Apr-2025].

[6] MSSQLTips, "SSIS Aggregate Transformation," MSSQLTips. [Online]. Available: <https://www.mssqltips.com/sqlservertip/3547/ssis-aggregate-transformation/>. [Accessed: 24-Apr-2025].

[7] Microsoft, "Sort Transformation - SQL Server Integration Services (SSIS)," Microsoft Learn. [Online]. Available: <https://learn.microsoft.com/en-us/sql/integration-services/data-flow/sort-transformation>. [Accessed: 24-Apr-2025].

[8] Microsoft, "Data Viewer - SQL Server Integration Services (SSIS)," Microsoft Learn. [Online]. Available: <https://learn.microsoft.com/en-us/sql/integration-services/data-flow/data-viewer>. [Accessed: 24-Apr-2025].

[9] Microsoft, "Multicast Transformation - SQL Server Integration Services (SSIS)," Microsoft Learn. [Online]. Available: <https://learn.microsoft.com/en-us/sql/integration-services/data-flow/multicast-transformation>. [Accessed: 24-Apr-2025].

[10] Microsoft, "Excel Destination - SQL Server Integration Services (SSIS)," Microsoft Learn. [Online]. Available: <https://learn.microsoft.com/en-us/sql/integration-services/data-flow/excel-destination>. [Accessed: 24-Apr-2025].

[11] Microsoft, "Flat File Destination - SQL Server Integration Services (SSIS)," Microsoft Learn. [Online]. Available: <https://learn.microsoft.com/en-us/sql/integration-services/data-flow/flat-file-destination>. [Accessed: 24-Apr-2025].

[12] Microsoft, "Configure the Windows Firewall to Allow SQL Server Access," Microsoft Learn. [Online]. Available: <https://learn.microsoft.com/en-us/sql/sql-server/install/configure-the-windows-firewall-to-allow-sql-server-access>. [Accessed: 24-Apr-2025].