

Institución Universitaria Digital de Antioquia

Ingeniería de Software y Datos

Bases de Datos II

Evidencia de aprendizaje 3. Creación de una base de datos de Staging y Proceso de transformación de datos

Curso: PREICA2401B010076 Grupo 15

Profesor: Víctor Hugo Mercado

Elaborado por: Juan Carlos Gutiérrez Riaño

Fecha: 23 de mayo de 2024

Índice

1. Introducción
2. Objetivos
3. Planteamiento del problema
4. Análisis del problema
5. Propuesta de la solución
6. Descripción del modelo estrella propuesto.
7. Diseño (Imagen) del modelo estrella donde se puedan observar las dimensiones, la tabla de hechos, sus campos, tipos de datos y relaciones.
8. Lista de dimensiones propuestas.
9. Dimensiones especificadas con columnas y qué tipos de datos van a almacenar.
10. Tabla de hechos, con sus campos y tipos de datos.
11. Proceso de transformación de datos
12. Conclusiones.
13. Webgrafía

1. Introducción

El presente trabajo plantea la solución hacia la consulta de una base de datos denominada Jardinería la cual podemos visualizar con SQL Server y poder resolver tres preguntas como lo son, identificar el producto más vendido, la categoría con más productos y el año con más ventas, por medio de un DataMart basado en el modelo estrella, después se utilizara el programa visual estudio community 2022 para hacer las practicas de creación de la base de datos jardinería y base de datos JCGR_STAGING-10076 y con estas realizar las consultas programadas con sus respectivos scripts .

2. Objetivos

- Determinar cuál son los elementos que nos permiten resolver las interrogantes de la mejor forma

- Determinar el producto más vendido, Determinar la categoría con más productos Y Determinar el año con más ventas con la metodología de estrella planteada.
- Determinar las relaciones entre la tabla de hecho y las dimensiones
- Determinar los datos requeridos en la tabla de hechos y dimensiones
- Obtener las consultas programadas de las dos bases de datos

3. Planteamiento del problema

Se requiere identificar las relaciones e identidades del modelo de la base datos Jardinería que cumpla con el requerimiento de:

- Poder determinar el producto más vendido
- Poder determinar la categoría con más productos
- Poder determinar el año con más ventas
- Verificar el funcionamiento adecuado del SQL Server con sus tablas y los resultados de las consultas con sus respectivos scripts

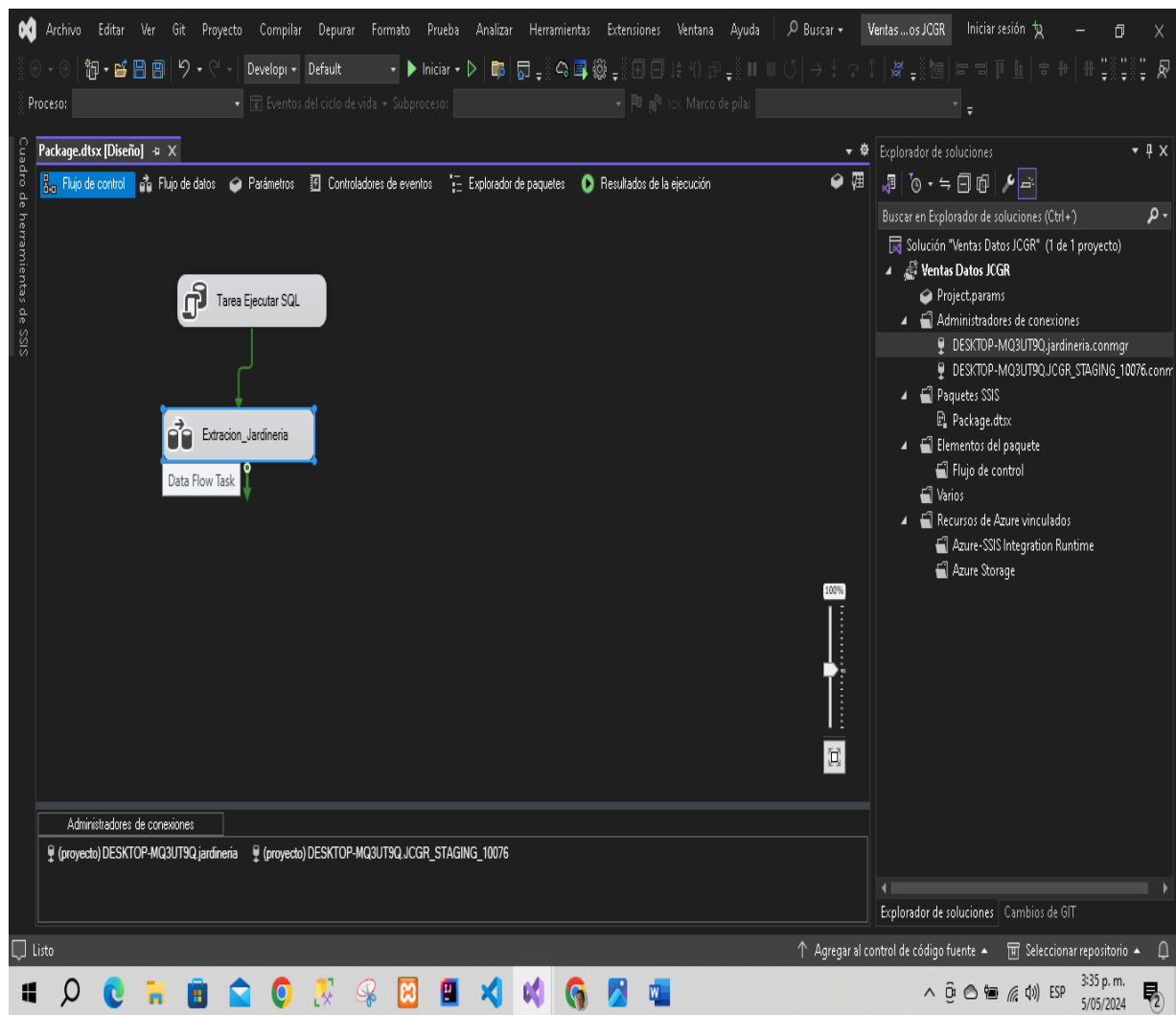
4. Análisis del problema

Para poder dar respuesta a las preguntas planteadas en el ejercicio, me dispongo a elaborar la tabla de hechos que es el punto de partida para la consulta de la información y después desarrollo el número de dimensiones necesarias para generar el DataMart a partir de un modelo en estrella contando con las tablas (pago, detalle, detalle de pedido, cliente, producto, categoría de producto, empleado, y oficina.)

Como segunda parte nos disponemos a crear el proyecto nuevo Ventas Datos JCGR junto a las dos bases de datos jardinería y base de datos JCGR_STAGING-10076,

- Flujo de control quedarían los siguientes componentes limpieza ejecutar SQL Y Extraccion_jardineria
- En la tercera parte aplicamos el proceso de transformación para las consultas analíticas , ayudandolos de las normalizaciones ,limpieza y enriquesimientos de los datos dados en jardineria junto a las consultas SQL que nos permite demostrar los cambios y consultas , aplicando las diferentes herramientas de la técnica ETL del . visual studio community.

- Flujo de datos con sus componentes donde se aprecian la base de datos de origen con sus respectivos SELEC (cliente, producto, tiempo, pedido) y los datos de destino con sus asignaciones (Tb cliente, Tb producto, Tb tiempo, Tb pedido)



Archivo Editar Ver Git Proyecto Compilar Depurar Formato Prueba Analizar Herramientas Extensiones Ventana Ayuda Buscar Ventas...os JCGR Iniciar sesión

Proceso: Eventos del ciclo de vida Subproceso: Marco de pila:

Package.dtsx [Diseño] X

Flujo de control Flujo de datos Parámetros Controladores de eventos Explorador de paquetes Resultados de la ejecución

Tarea Flujo de datos: Extraccion_Jardineria

Origen_Cliente Destino_TbCliente

Origen_Producto Destino_TbProducto

Origen_Tiempo Destino_TbTiempo

Origen_Pedido Destino_TbPedido

Administradores de conexiones

(proyecto) DESKTOP-MQ3UT9Q.jardineria (proyecto) DESKTOP-MQ3UT9Q.JCGR_STAGING_10076

Explorador de soluciones

Buscar en Explorador de soluciones (Ctrl+)

Solución "Ventas Datos JCGR" (1 de 1 proyecto)

- Ventas Datos JCGR
 - Project.params
 - Administradores de conexiones
 - DESKTOP-MQ3UT9Q.jardineria.connmgr
 - DESKTOP-MQ3UT9Q.JCGR_STAGING_10076.connr
 - Paquetes SSIS
 - Package.dtsx
 - Elementos del paquete
 - Flujo de control
 - Varios
 - Recursos de Azure vinculados
 - Azure-SSIS Integration Runtime
 - Azure Storage

Explorador de soluciones Cambios de GIT

Lista

Agregar al control de código fuente Seleccionar repositorio

3:38 p.m. 5/05/2024

- SQL Server con sus tablas creadas en JCGR_STAGING-10076 y los resultados de las consultas con sus selec con el contenido específico de los scripts

The screenshot displays the Microsoft SQL Server Management Studio (SSMS) interface. The left pane shows the Object Explorer with the database 'JCGR_STAGING_10076' selected. The central pane shows a SQL query with several statements. The bottom pane shows the results of the query, which is a table with three columns: 'dia', 'mes', and 'año'. The results are as follows:

	dia	mes	año
1	1	4	2009
2	1	8	2008
3	1	10	2008
4	2	1	2009
5	2	2	2009
6	2	3	2009
7	3	1	2009

The status bar at the bottom indicates that the query was executed successfully and that there are 78 rows in the results.

SQLQuery5.sql - DESKTOP-MQ3UT9Q\jardineria (DESKTOP-MQ3UT9Q\jcgut (70))* - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

Execute

SQLQuery5.sql - DE...Q3UT9Q\jcgut (70)* DESKTOP-MQ3UT9Q\j...eria - dbo.View_1* SQLQuery3.sql - DE...Q3UT9Q\jcgut (52)*

```
--Cliente
SELECT ID_cliente, nombre_cliente, ciudad, region, pais from cliente;

--Tiempo
Select DAY(fecha_pedido) as dia, MONTH(fecha_pedido) as mes, YEAR(fecha_pedido) as año from pedido;

--Producto
SELECT P.ID_producto, P.precio_venta, P.CodigoProducto, P.nombre, C.Desc_Categoria from producto P
inner join Categoria_producto C on P.Categoria = C.Id_Categoria;

--Pedido
Select Dp.ID_detalle_pedido, Dp.ID_producto, PE.ID_cliente, Dp.cantidad, PE.fecha_entrega from pedido Pe
inner join detalle_pedido Dp on Pe.ID_pedido = Dp.ID_pedido;
```

110 %

Results Messages

	dia	mes	año
1	1	4	2009
2	1	8	2008
3	1	10	2008
4	2	1	2009
5	2	2	2009
6	2	3	2009

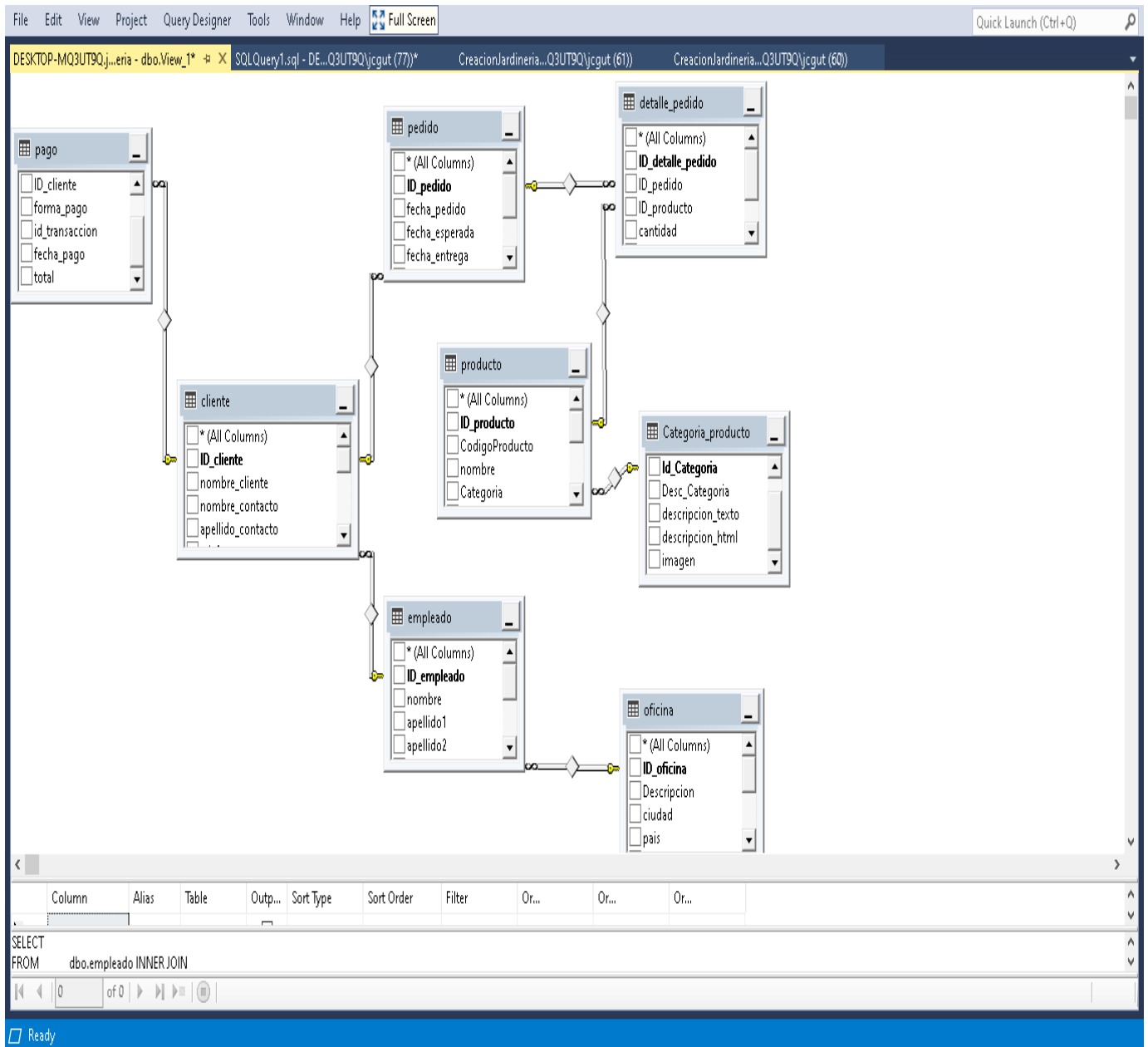
Query executed successfully.

DESKTOP-MQ3UT9Q (16.0 RTM) | DESKTOP-MQ3UT9Q\jcgut ... | jardineria | 00:00:00 | 78 rows

Ready Ln 13 Col 61 Ch 61 INS

4:13 p. m.
5/05/2024

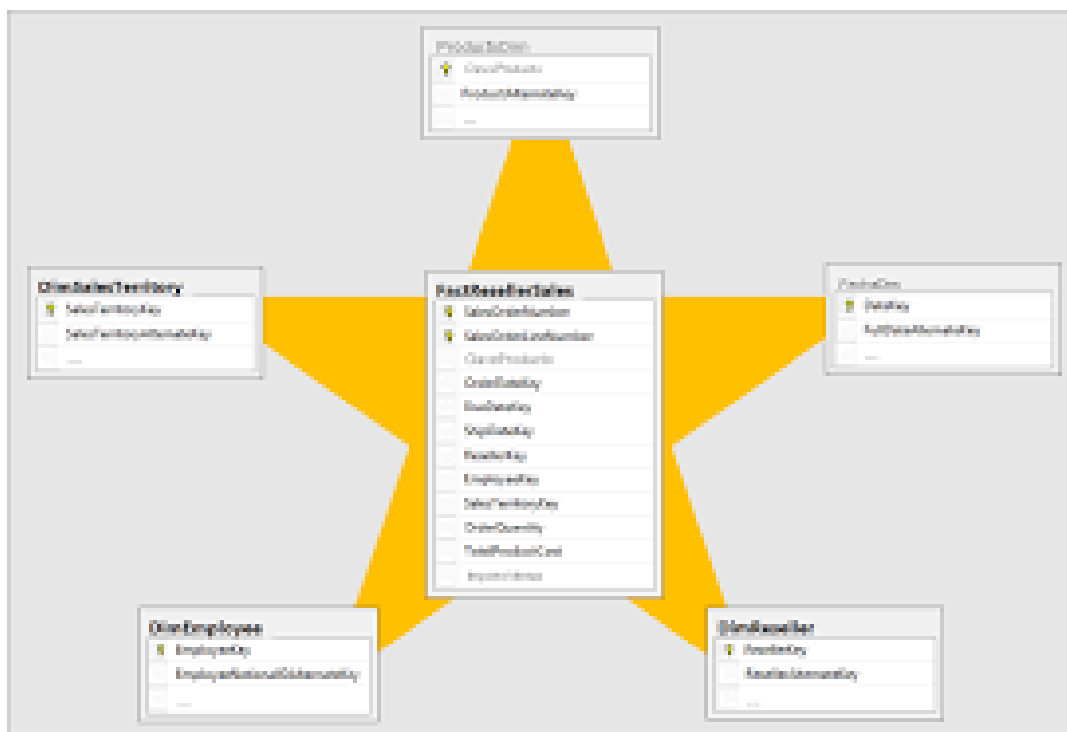
Tablas de la base de datos de jardinería



5. Propuesta de la solución

La propuesta es elaborar el método estrella partiendo del análisis de los datos que se pueden visualizar en la base de datos de Jardinería, en esta se aprecian propiedades con datos relevantes para el ejercicio

6. Descripción del modelo estrella propuesto.



Ejemplo modelo estrella

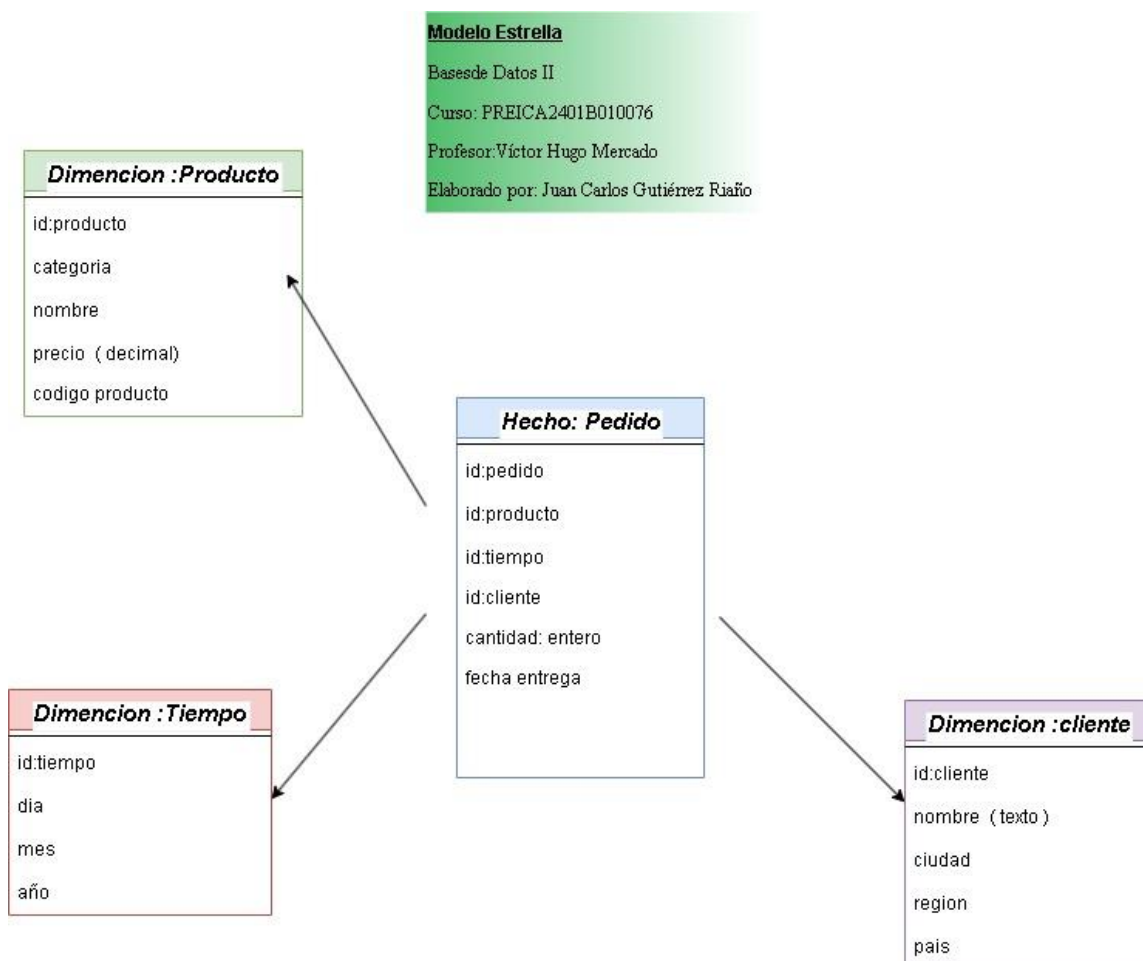
<https://learn.microsoft.com/es-es/power-bi/guidance/star-schema>

La propuesta es elaborar el método estrella partiendo de una tabla de hechos llamada pedidos junto a sus tipos de datos y rodeada de cuatro dimensiones llamadas: dimensión producto, dimensión categoría, dimensión tiempo y dimensión cliente.

7. Diseño (Imagen) del modelo estrella donde se puedan observar las dimensiones, la tabla de hechos, sus campos, tipos de datos y relaciones.

Vinculo de la imagen:

<https://app.diagrams.net/#G1xOiPeoNoeBIYDDS1nnaYbMbUxNcHXjCG#%7B%22pageld%22%3A%22C5Rbs43oDa-KdzZeNtuy%22%7D>



8. Lista de dimensiones propuestas.

Las listas de dimensiones serán:

- Dimensión de producto
- Dimensión de cliente
- Dimensión de tiempo
- Dimensión de pedido

9. Dimensiones especificadas con columnas y qué tipos de datos van a almacenar.

Dimensión Producto	
Id: producto	entero
Nombre producto	string
precio	decimal
categoría	string

Código producto	string
-----------------	---------------

Dimensión Cliente	
Id: cliente	entero
Nombre	string
Ciudad	string
Región	string
País	string

Dimensión Tiempo	
Id: tiempo	entero
Día	entero
Mes	entero

Año	entero
-----	--------

10. Tabla de hechos, con sus campos y tipos de datos.

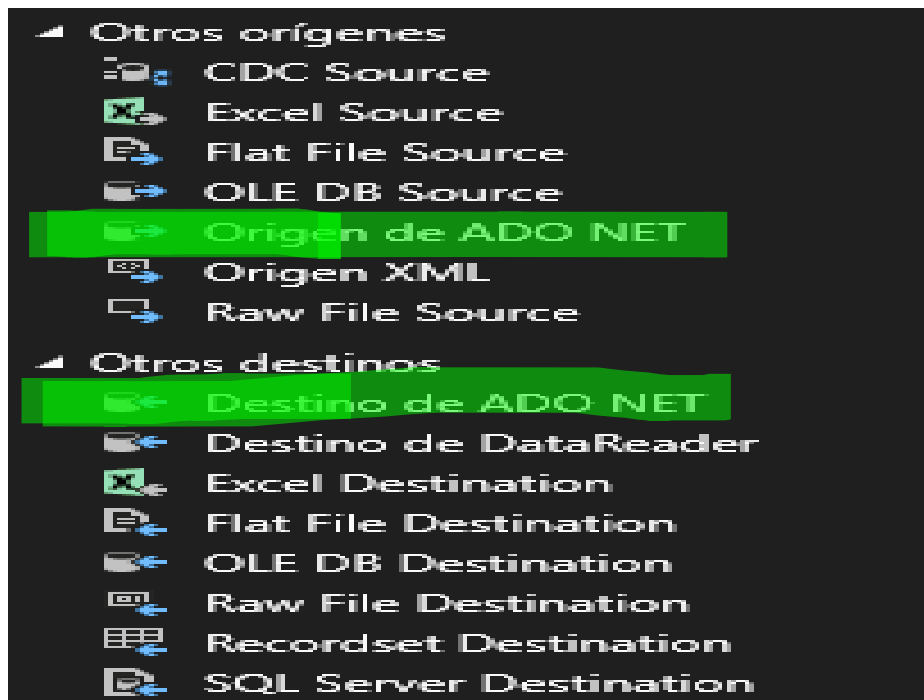
Tabla de Hechos: Pedido	
Id: pedido	entero
Id: producto	entero
Id: Tiempo	entero
Id: cliente	entero
cantidad	entero
Fecha entrega	entero

11. Proceso de transformación de datos

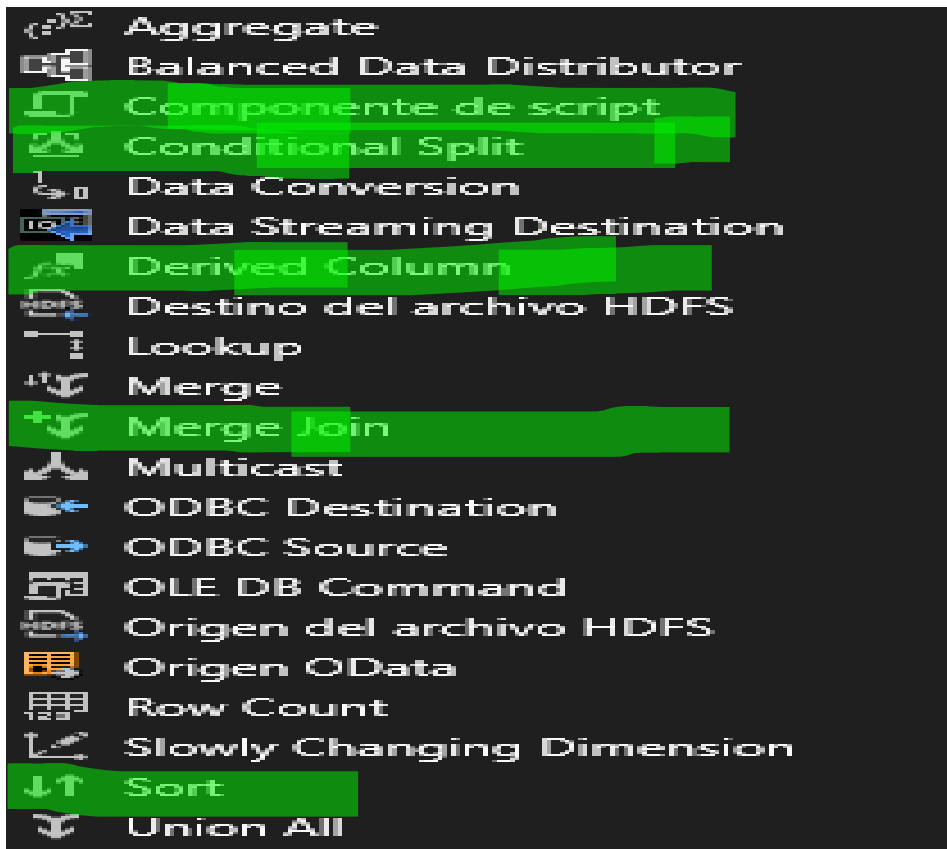
Para desarrollar la transformación y verificar su funcionamiento utilizamos las diferentes herramientas de la técnica ETL, visual studio community, también pudimos cruzar datos, orígenes ,destinos,



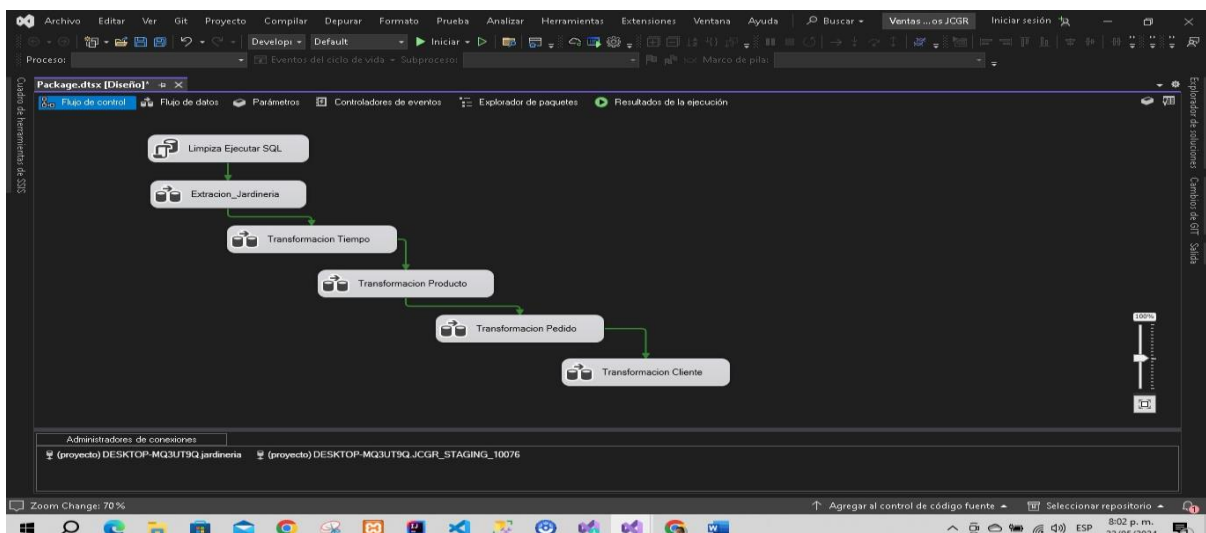
flujo de control



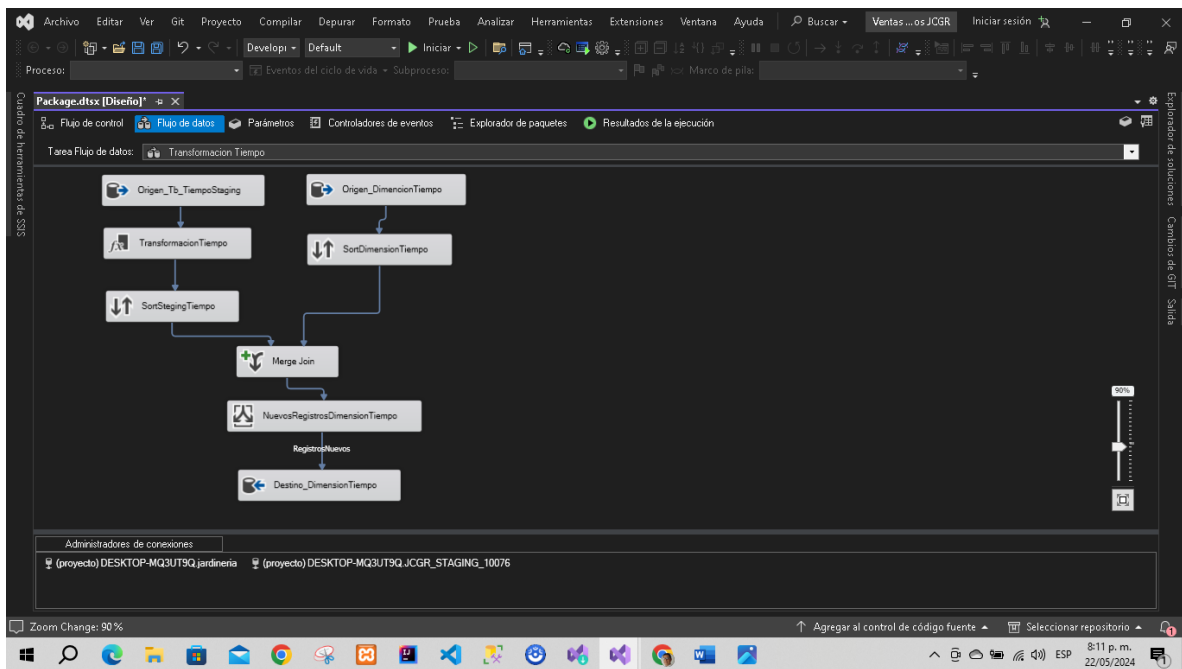
origen y destino ADO NET



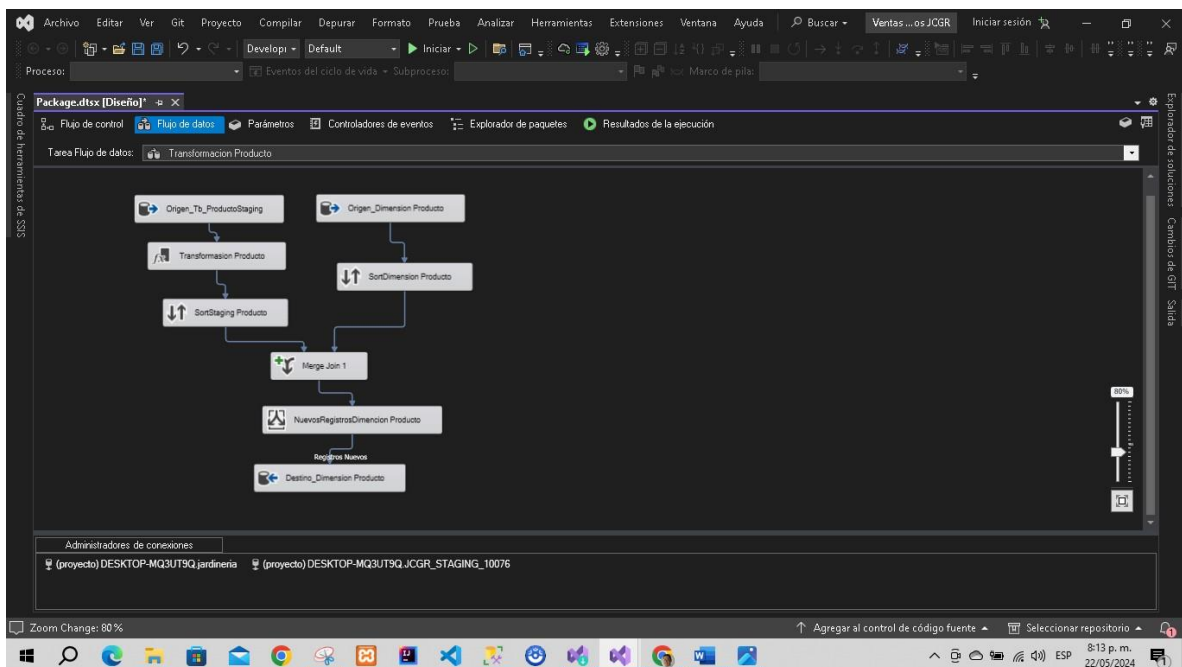
Visualización del esquema de limpieza, extracción, transformaciones de tiempo, producto, pedido y cliente



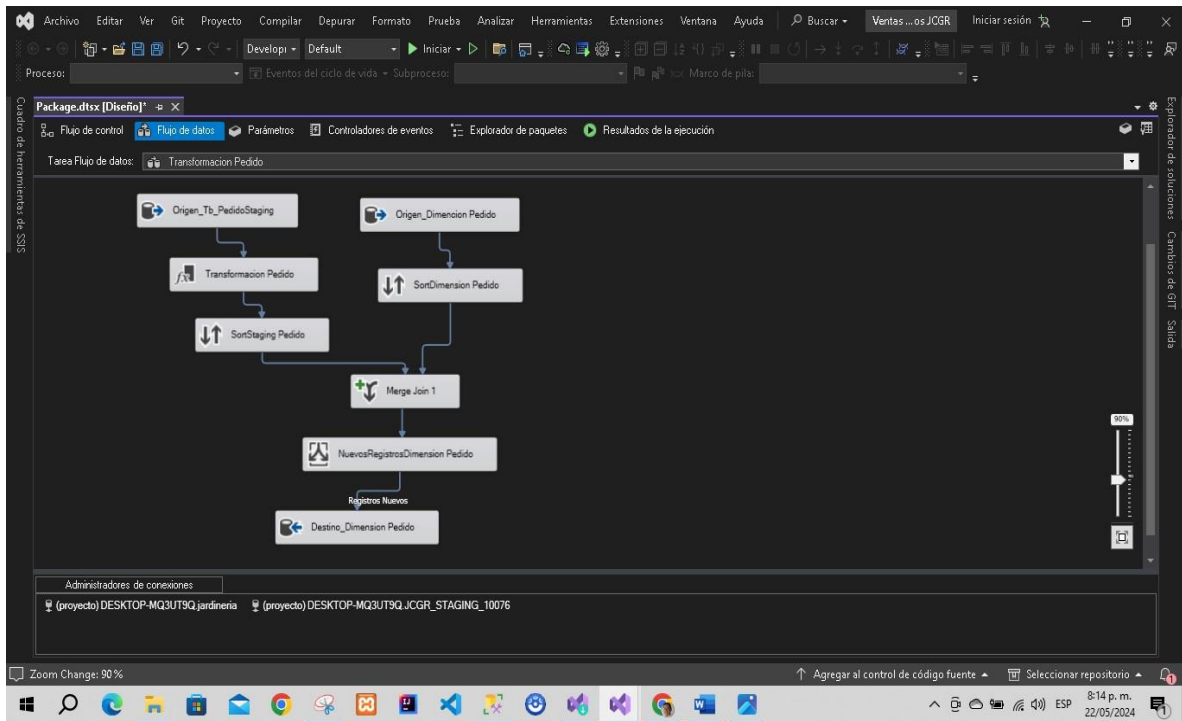
Transformación tiempo



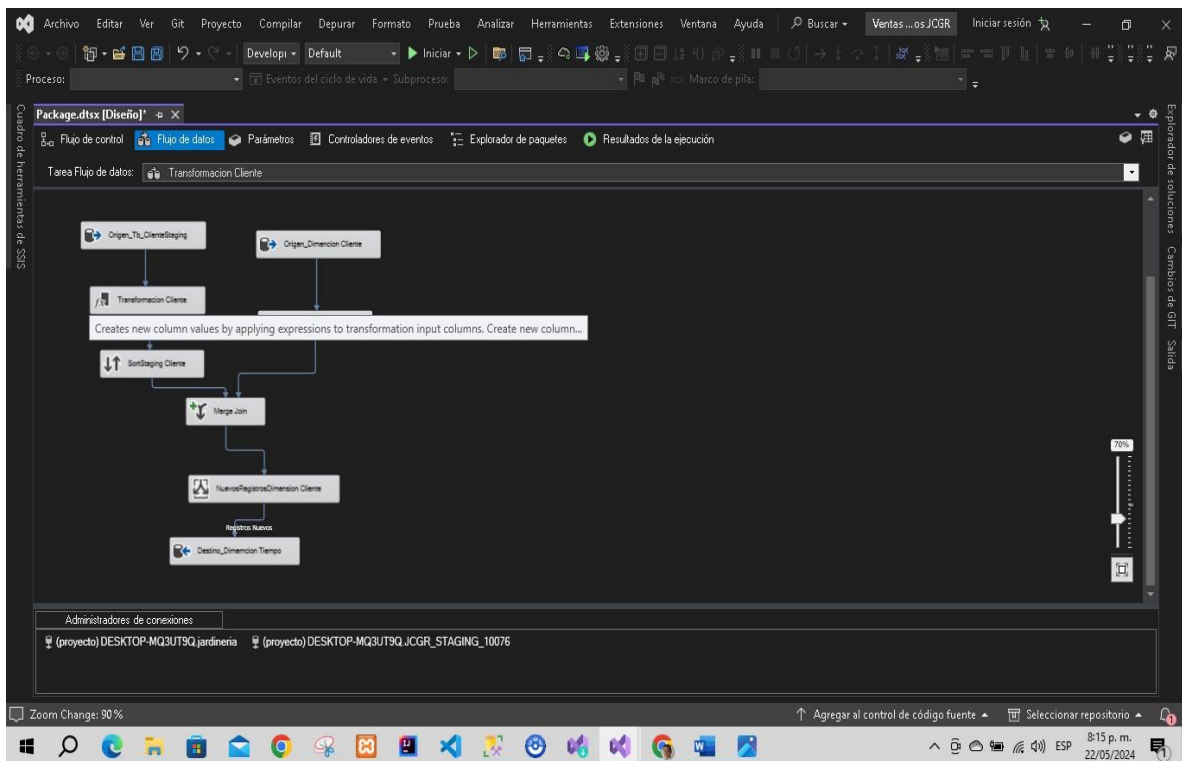
Transformación producto



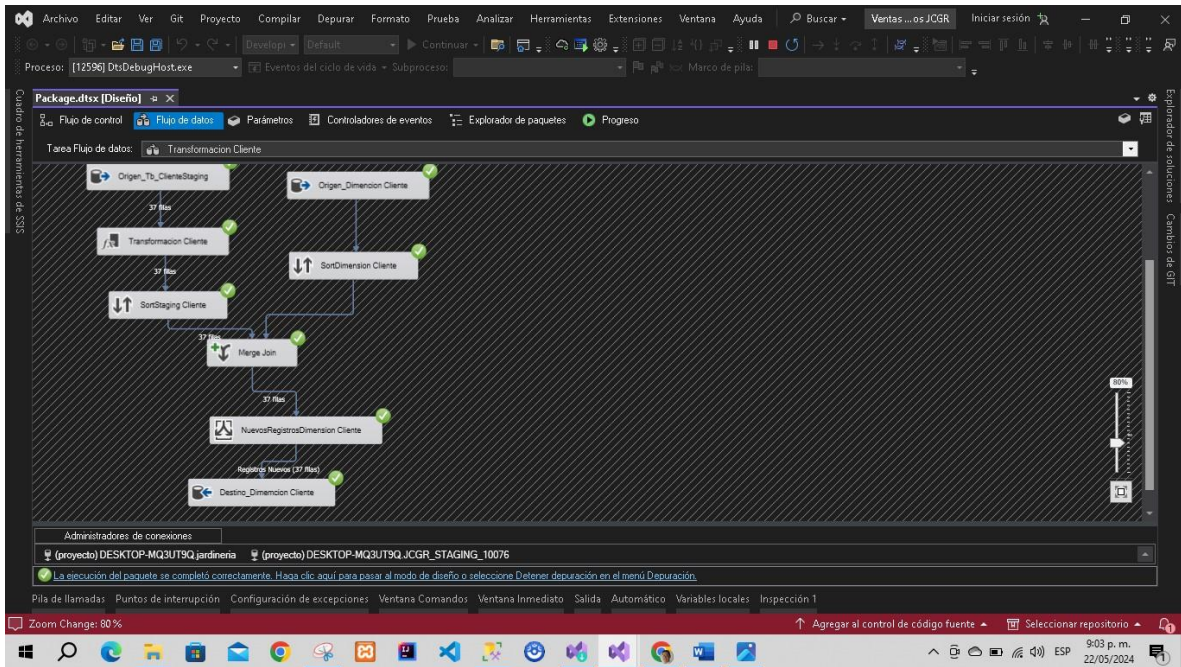
Transformación pedido



Transformación cliente



- Demostración de enriquecimiento con tabla cliente con 37 resultados, se insertará un nuevo registro para que muestre 38 resultados



DESKTOP-MQ3UT9Q.jardineria - dbo.cliente - Microsoft SQL Server Enterprise Manager

Object Explorer

Connect +

DESKTOP-MQ3UT9Q.jardineria - dbo.cliente

ID_cliente	nombre_cliente	nombre_conta...	apellido_conta...	telefono	fax	linea_direccion1	linea_direccion2	ciudad	re...
17	Naturajardin	Victoria	Cruz	612343529	916548735	Plaza Magallón ...	NULL	Madrid	re...
18	Golf S.A.	Luis	Martinez	916458762	912354475	C/Estancado	NULL	Santa cruz de T...	loli
19	Americh Golf ...	Mario	Suarez	964493072	964493063	C/Letardo	NULL	Barcelona	Ca
20	Aloha	Cristian	Rodriguez	916485852	914489898	C/Roman 3	NULL	Canarias	Ca
21	El Prat	Francisco	Camacho	916882323	916493211	Avenida Tibidabo	NULL	Barcelona	Ca
22	Sotogrande	Maria	Santillana	915576622	914825645	C/Paseo del Par...	NULL	Sotogrande	Ca
23	Vivero Humanes	Federico	Gomez	654987690	916040875	C/Miguel Eche...	NULL	Humanes	Mt
24	Fuenla City	Tony	Muñoz Mena	675942139	915483754	C/Callo 52	NULL	Fuenlabrada	Mt
25	Jardines y Mans...	Eva Maria	Sánchez	916877445	914477777	Poligono Indus...	NULL	Madrid	Mt
26	Jardinerias Matí...	Matías	San Martín	916544147	917897474	C/Francisco Ar...	NULL	Madrid	Mt
27	Agrojardin	Benito	Lopez	675432926	916549264	C/Mar Caspio 43	NULL	Getafe	Mt
28	Top Campo	Joseluis	Sanchez	685746512	974315924	C/Ibiza 32	NULL	Humanes	Mt
29	Jardineria Sara	Sara	Marquez	675124537	912475843	C/Lima 1	NULL	Fuenlabrada	Mt
30	Campohermoso	Luis	Jimenez	645925376	916159116	C/Peru 78	NULL	Fuenlabrada	Mt
31	france telecom	Fraçois	Toulou	(33)5120578961	(33)5120578961	6 place d Allera...	NULL	Paris	NL
32	Musée du Louvre	Pierre	Delacroix	(33)0140205050	(33)0140205442	Quai du Louvre	NULL	Paris	NL
33	Totfruti S.A	Jacob	Jones	2 9261-2433	2 9283-1695	level 24, St. Mar...	NULL	Sydney	Nu
34	Flores S.L.	Antonio	Romero	654325981	685249700	Avenida España	NULL	Madrid	Fu
35	The Magic Gard...	Richard	Mcain	926523468	9364875882	Lighting Park	NULL	London	Loi
36	El Jardin Vivien...	Justin	Smith	2 8005-7161	2 8005-7162	176 Cumberland...	NULL	Sydney	Nu
38	carlos	pablo	rios	78654	73567	calle	camera	zipa	me
38	juan	rios	rios	4543234	786543	calle	sur	sopo	sur
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NL

38 of 38 | Cell is Modified...

Archivo Editar Ver Gits Proyects Compilar Depurar Formato Prueba Analizar Extensiones Ventana Ayuda Buscar Ventas...os JOGR Iniciar sesión

Proceso: [11032] DtsDebugHost.exe Eventos del ciclo de vida Subprocesos: Marco de pila:

Package.dtsx [Diseño]

Flujo de control Flujo de datos Parámetros Controladores de eventos Explorador de paquetes Progreso

Tarea Flujo de datos: Transformación Cliente

The diagram illustrates a data flow task named 'Transformación Cliente'. It starts with two source components: 'Origen_Tb_ClienteStaging' and 'Origen_Dimension Cliente'. 'Origen_Tb_ClienteStaging' flows through a 'Transformacion Cliente' task and a 'SortStaging Cliente' sort task. 'Origen_Dimension Cliente' flows through a 'SortDimension Cliente' sort task. Both sort tasks feed into a 'Merge Join' component. The output of the 'Merge Join' flows into a 'NuevosRegistrosDimension Cliente' task, which then feeds into a 'Registros Nuevos (1 Row)' task, and finally into the 'Destino_Dimension Cliente' destination. Data flow rates are indicated on the connections: 38 Rows, 37 Rows, 38 Rows, 37 Rows, 14 Rows, 38 Rows, and 1 Row. Two blue arrows point to the 'Merge Join' and 'NuevosRegistrosDimension Cliente' components. The task is marked as 'Completed' with a green checkmark.

Administradores de conexiones

(proyecto) DESKTOP-MQ3UT9Q.jardineria (proyecto) DESKTOP-MQ3UT9Q.JOGR_STAGING_10076

La ejecución del paquete se completó correctamente. Haga clic aquí para pasar al modo de diseño o seleccione [Detener depuración](#) en el menú [Depuración](#).

Pila de llamadas: Puntos de interrupción Configuración de excepciones Ventana Comandos Ventana Inmediato Salida Automático Variables locales Inspección 1

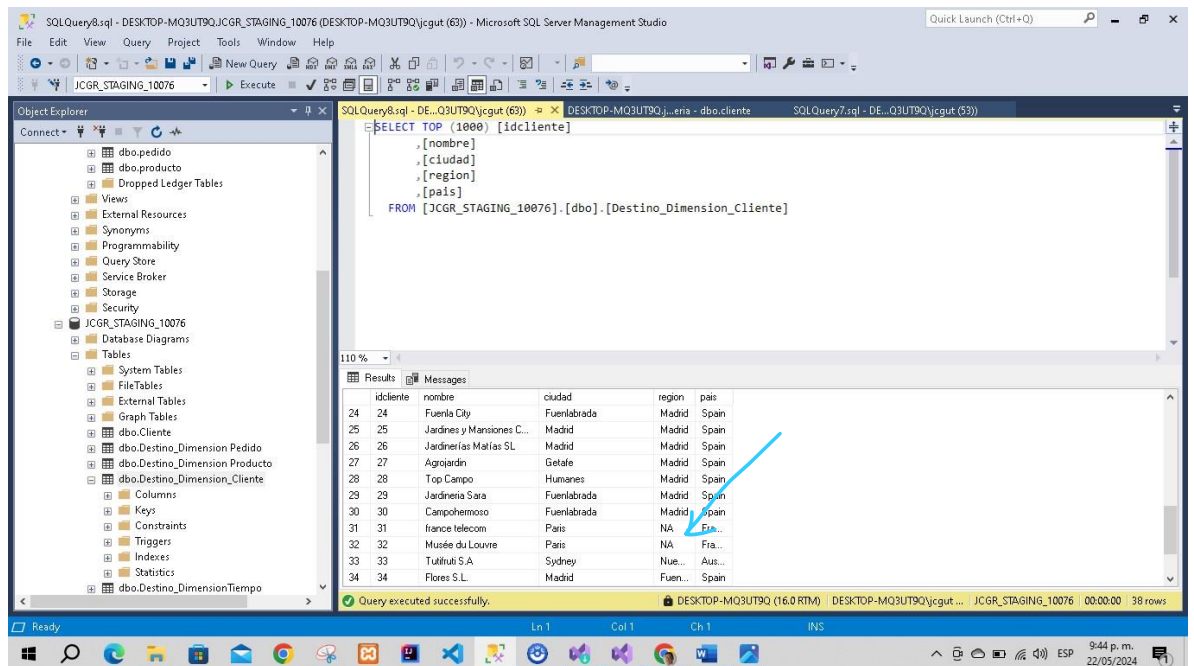
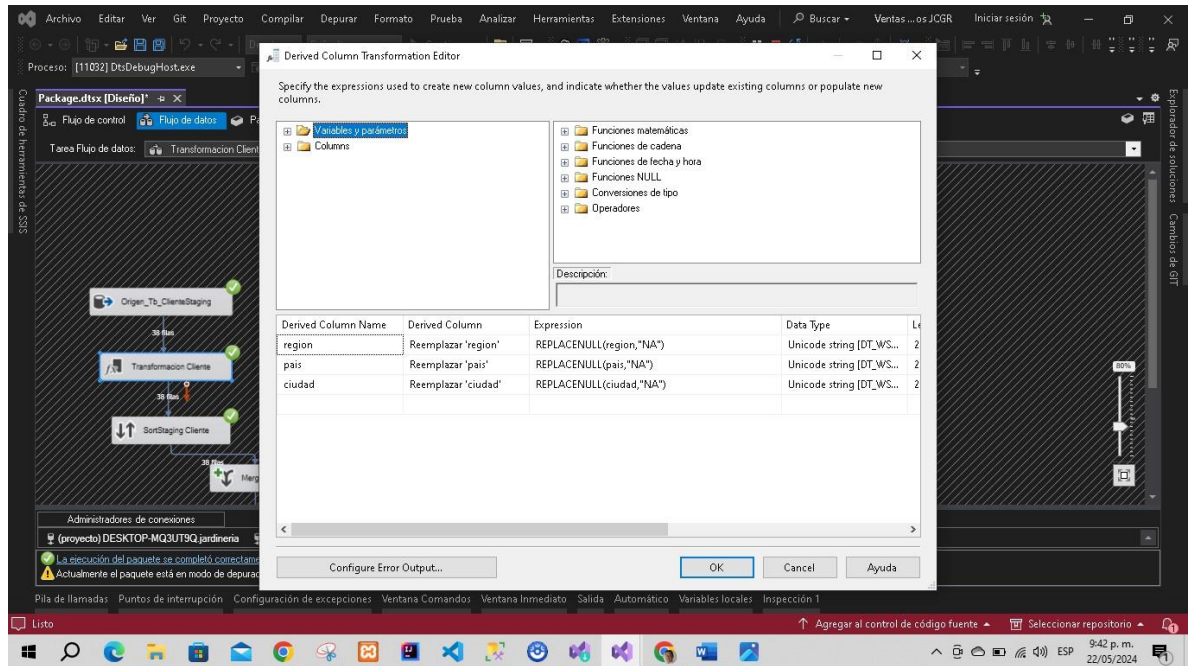
Listo

Agregar al control de código fuente Seleccionar repositorio

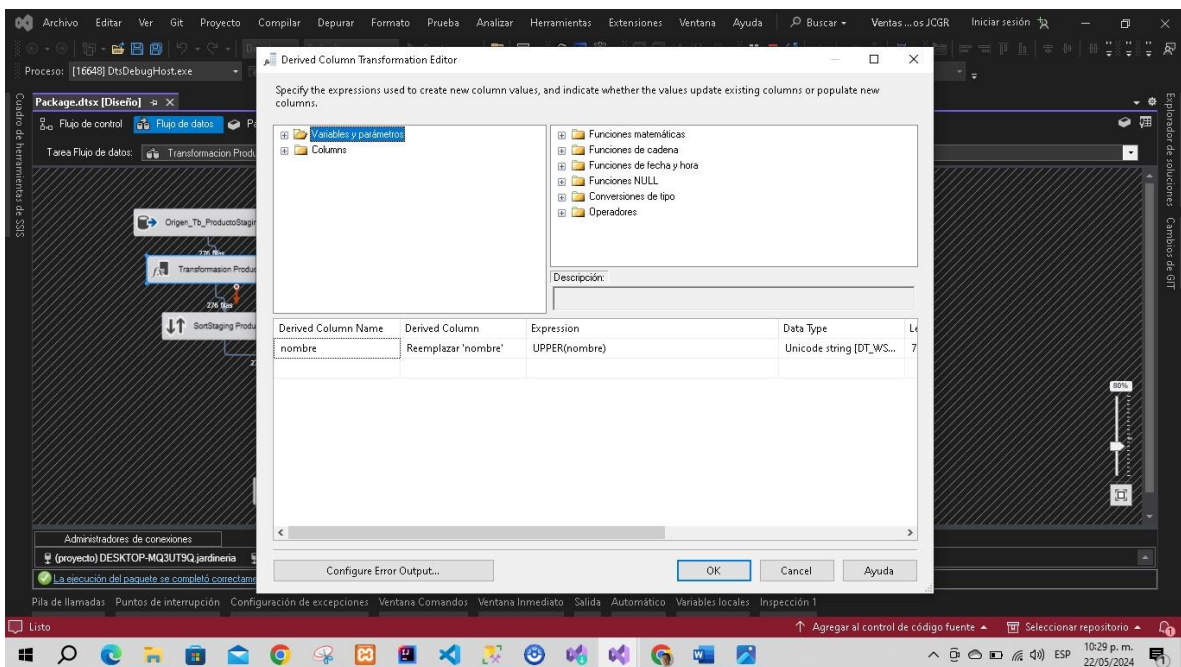
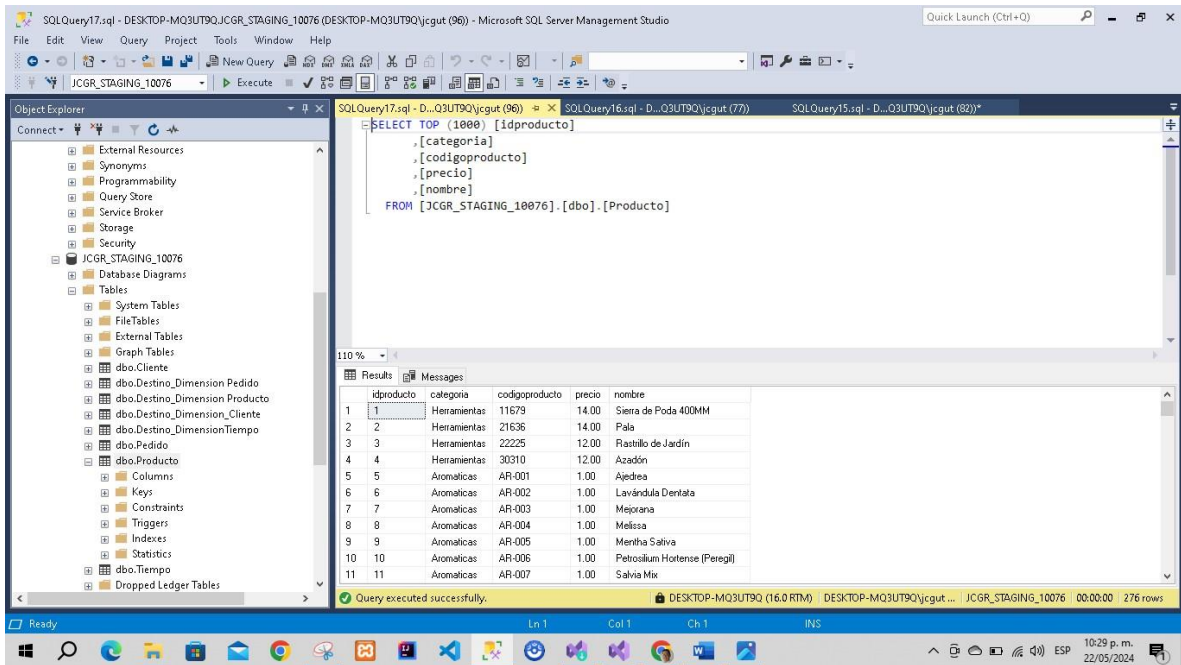
9-12 p.m. 22/05/2024

- Demostración de transformación con país, región y ciudad se cambia Null por NA resultados,

[illegible]



- Demostración de transformación cambio de nombre de producto que está en minúscula a mayúsculas



SQLQuery16.sql - DESKTOP-MQ3UT9Q\JCGR_STAGING_10076 (DESKTOP-MQ3UT9Q\jcgut (77)) - Microsoft SQL Server Management Studio

Object Explorer

- External Resources
- Synonyms
- Programmability
- Query Store
- Service Broker
- Storage
- Security
- JCGR_STAGING_10076
 - Database Diagrams
 - Tables
 - System Tables
 - FileTables
 - External Tables
 - Graph Tables
 - dbo.Cliente
 - dbo.Destino_Dimension Pedido
 - dbo.Destino_Dimension Producto
 - dbo.Destino_Dimension_Cliente
 - dbo.Destino_DimensionTiempo
 - dbo.Pedido
 - dbo.Producto
 - dbo.Tiempo
 - Dropped Ledger Tables
 - Views
 - External Resources
 - Synonyms
 - Programmability
 - Query Store
 - Service Broker

SQLQuery16.sql - D:\MQ3UT9Q\jcgut (77)

```
SELECT TOP (1000) [Idproducto]
,[categoria]
,[codigoproducto]
,[precio]
,[nombre]
FROM [JCGR_STAGING_10076].[dbo].[Destino_Dimension Producto]
```

Results

Idproducto	categoria	codigoproducto	precio	nombre
1	Herramientas	11679	14.00	SIERRA DE PODA 400MM
2	Herramientas	21636	14.00	PALA
3	Herramientas	22225	12.00	RASTRILLO DE JARDIN
4	Herramientas	30310	12.00	AZADÓN
5	Aromáticas	AR-001	1.00	AJEDREA
6	Aromáticas	AR-002	1.00	LAVANDULA DENTATA
7	Aromáticas	AR-003	1.00	MEJORANA
8	Aromáticas	AR-004	1.00	MELISSA
9	Aromáticas	AR-005	1.00	MENTHA SATIVA
10	Aromáticas	AR-006	1.00	PETROSILIMUM HORTENSE (PEREGIL)
11	Aromáticas	AR-007	1.00	SALVIA MDX

Query executed successfully. DESKTOP-MQ3UT9Q (16.0 RTM) DESKTOP-MQ3UT9Q\jcgut ... JCGR_STAGING_10076 00:00:00 276 rows

Archivo Editar Ver Git Proyecto Compilar Depurar Formato Prueba Analizar Herramientas Extensiones Ventana Ayuda Buscar Ventas...os JCGR Iniciar sesión

Proceso: [16648] DtsDebugHost.exe Eventos del ciclo de vida Subprocesos

Package.dtsx [Diseño]

Flujo de control Flujo de datos Parámetros Controladores de eventos Explorador de paquetes Progreso

Calidad de herramientas de SSIS

Diagrama de flujo de datos:

```
graph TD
    A[Limpieza Ejecutar SQL] --> B[Extracion_jardineria]
    B --> C[Transformacion Tiempo]
    C --> D[Transformacion Producto]
    D --> E[Transformacion Pedido]
    E --> F[Transformacion Cliente]
```

Administradores de conexiones

- (proyecto) DESKTOP-MQ3UT9Q\jardineria
- (proyecto) DESKTOP-MQ3UT9Q\JCGR_STAGING_10076

La ejecución del paquete se completó correctamente. Haga clic aquí para pasar al modo de diseño o seleccione Opciones de depuración en el menú Depuración.

Pila de llamadas Puntos de interrupción Configuración de excepciones Ventana Comando Ventana Inmediata Salida Automático Variables locales Inspección 1

Listo Agregar al control de código fuente Seleccionar repositorio

12. Conclusiones.

Por medio de este ejercicio se comprendió la importancia que tiene el modelo de estrella para agilizar y optimizar los registros, también las características de la tabla de hechos que discrimina los datos cuantitativos de la base de datos junto a sus dimensiones las cuales describen los detalles (eventos) de la tabla de hechos.

Aprendí que con estos componentes realizaremos las consultas pertinentes, ejecutando los scripts necesarios para trabajar el Data Mart en una unidad de negocio con datos normalizados para así poder hacer la elaboración de las dos bases de datos (jardinería y JCGR_STAGING-10076) en los programas visual estudio community 2022 y SQL server., también aprendí a hacer cruce de datos, transformaciones, insertar filas, consultas, filtrar los Null., con las herramienta ETL.

13. Webgrafía

<https://uctunexpo.autanabooks.com/index.php/uct/article/view/342>

H Zepa, R Garcia, H Izquierdo - Universidad ..., 2020 - uctunexpo.autanabooks.com

<http://ingenieriasimple.com/conred/el%20modelo%20relacional.pdf>

J Quiroz - Boletín de Política Informática, 2003 - ingenieriasimple.com

<https://repositorio.utp.edu.pe/handle/20.500.12867/2101>

ME Dianderas Alcántara - 2019 - repositorio.utp.edu.pe

<https://learn.microsoft.com/es-es/power-bi/guidance/star-schema>

Descripción de un esquema de estrella e importancia para Power BI

Artículo

22/03/2023

9 colaboradores

<https://www.youtube.com/watch?v=JOY868aVlJw>

Creación de una Tabla de Hechos @dathalearning8063