Proyecto integrador-Repositorio de todas las actividades

Jose Quevin Ordoñez Sanchez

2025

.

Institución Universitaria Digital de Antioquia

Medellín, Antioquia.

Bases de Datos II

Victor Hugo Mercado

# Introducción

# En el contexto de la inteligencia de negocios y el análisis de datos, la calidad y la estructuración de la información son fundamentales para garantizar resultados confiables y procesables. Este documento describe el proceso de Extracción, Transformación y Carga (ETL) diseñado para migrar datos desde fuentes operacionales hacia un Data Mart basado en el modelo estrella, optimizado para consultas analíticas y reporting.

# El objetivo principal de este proceso es asegurar que los datos crudos sean depurados, normalizados y enriquecidos según los requisitos del negocio, antes de ser cargados en las tablas de dimensiones y hechos del Data Mart. Para lograrlo, se aplicaron técnicas de limpieza de datos (manejo de valores nulos, corrección de formatos, eliminación de duplicados), normalización (estandarización de unidades, conversión de tipos de datos) y enriquecimiento (generación de campos calculados, categorizaciones y agregaciones).

# La transformación se implementó mediante consultas SQL y herramientas de ETL, asegurando que los datos cumplan con los estándares de calidad antes de su carga final. Posteriormente, se diseñaron scripts de inserción para poblar las tablas del Data Mart, verificando la integridad referencial y la consistencia de los datos.

# Este informe detalla cada etapa del proceso, incluyendo:

# Las técnicas de transformación aplicadas y su justificación.

# Las consultas SQL y flujos ETL utilizados.

# El mecanismo de carga en el modelo estrella.

# Las validaciones realizadas para garantizar la precisión de los datos.

# Además, se incluyen referencias al modelo dimensional implementado, explicando cómo las tablas de hechos y dimensiones se relacionan para facilitar el análisis multidimensional. El resultado es un Data Mart confiable y optimizado que servirá como base para la generación de reportes, dashboards y análisis avanzados.

# Planteamiento del problema

# A partir del análisis de la base de datos Jardinería, se han identificado diversas dificultades en su administración, principalmente debido a la falta de estandarización de los datos, la duplicidad de registros y errores en la captura de información. Estas deficiencias afectan directamente la precisión de los reportes y comprometen la toma de decisiones basada en datos.

# Asimismo, la ausencia de una capa intermedia para la limpieza y validación de la información antes de su análisis ha generado inconsistencias en los reportes gerenciales y dificultades en la integración con otros sistemas. Para abordar estos problemas, se propone la implementación de una base de datos Staging, que optimice la organización de los datos, garantice su confiabilidad y facilite futuras consultas y análisis.

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

Diagrama

El contenido generado por IA puede ser incorrecto.

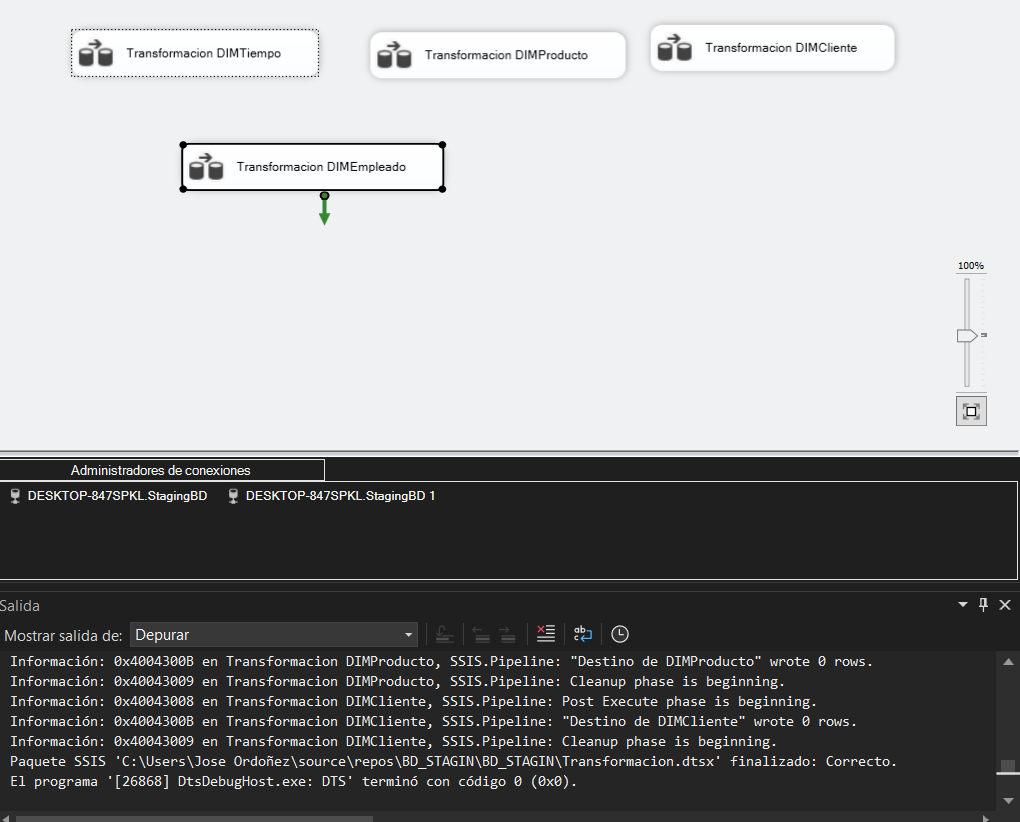
Interfaz de usuario gráfica

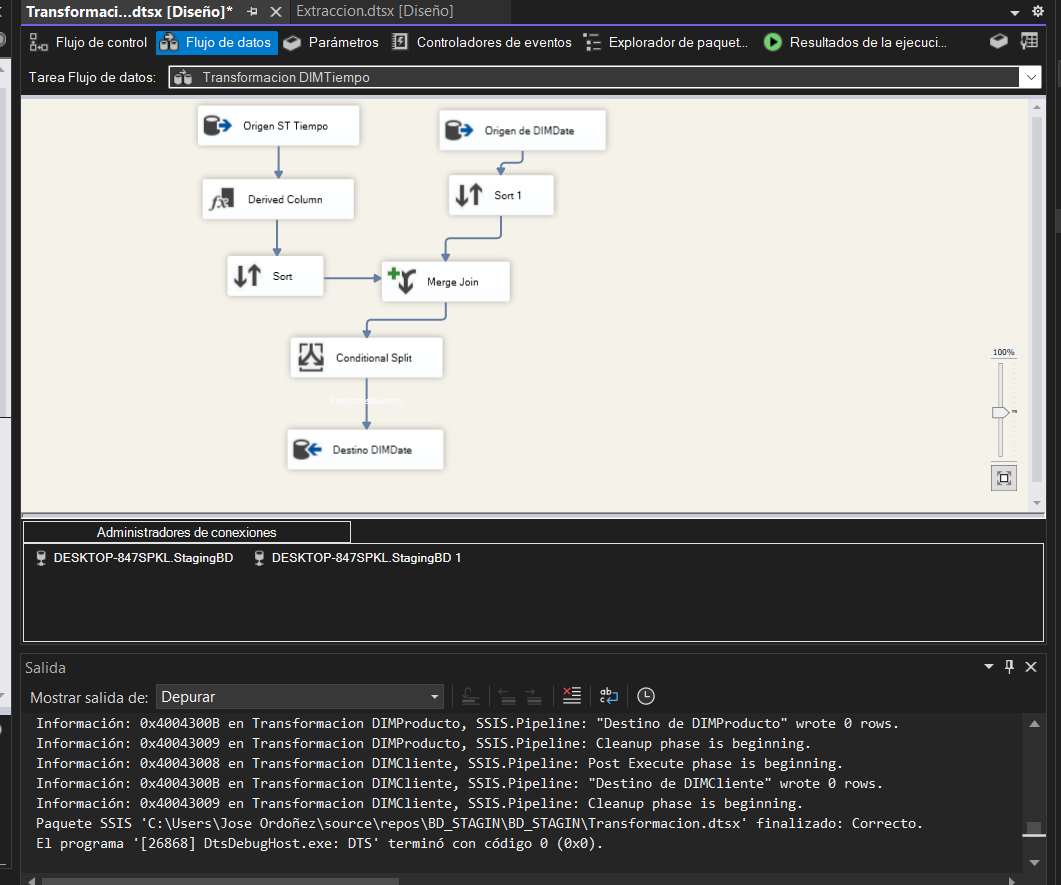
El contenido generado por IA puede ser incorrecto.

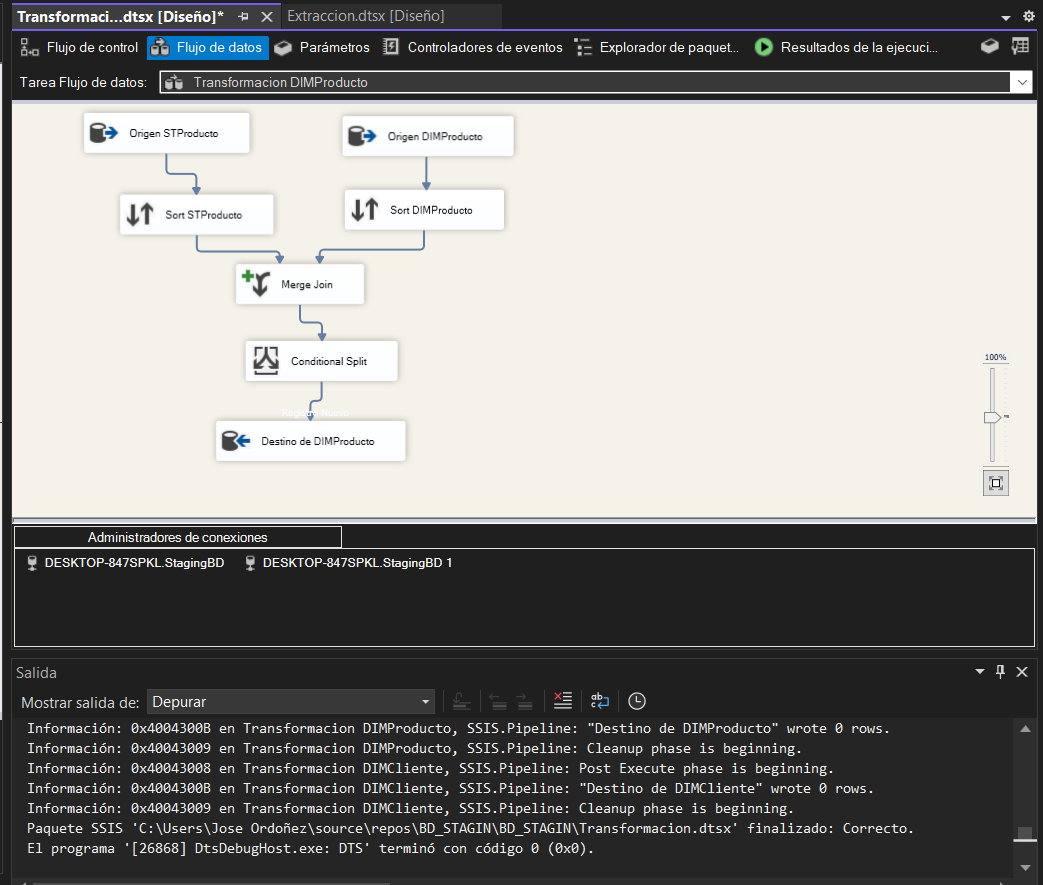
Una captura de pantalla de una computadora

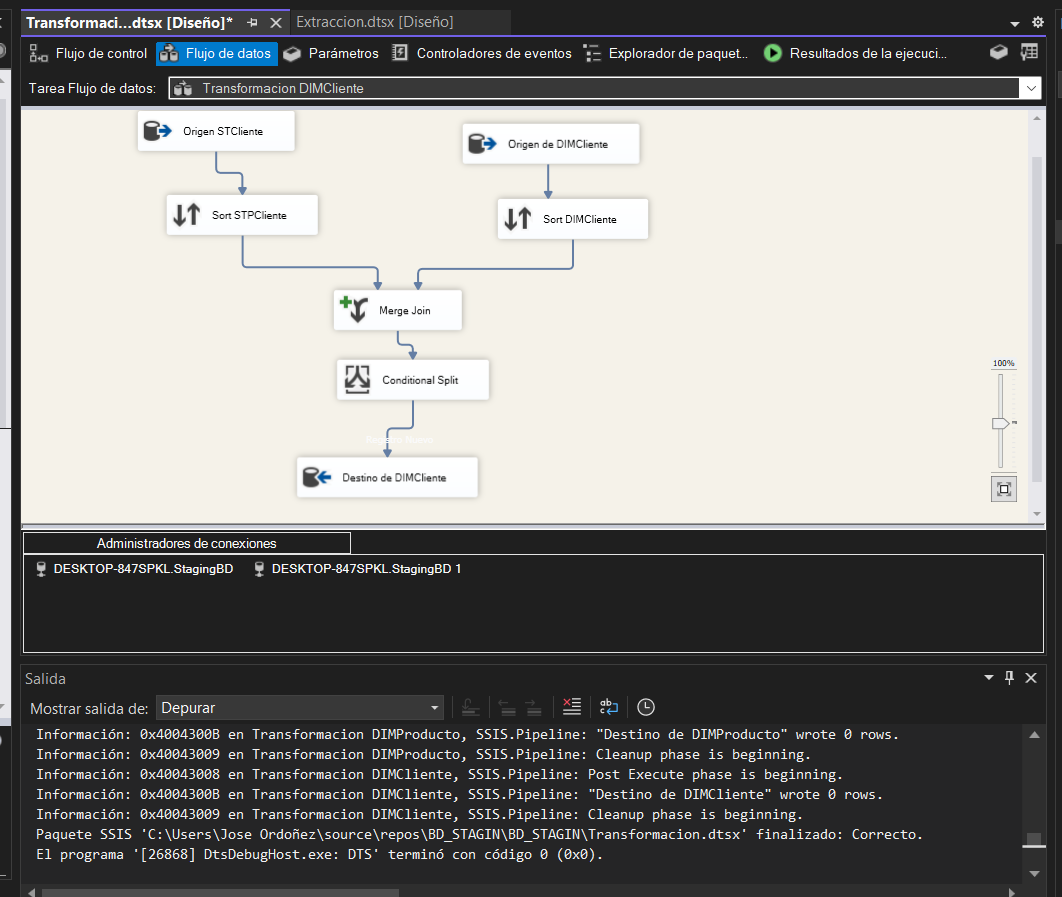
El contenido generado por IA puede ser incorrecto.

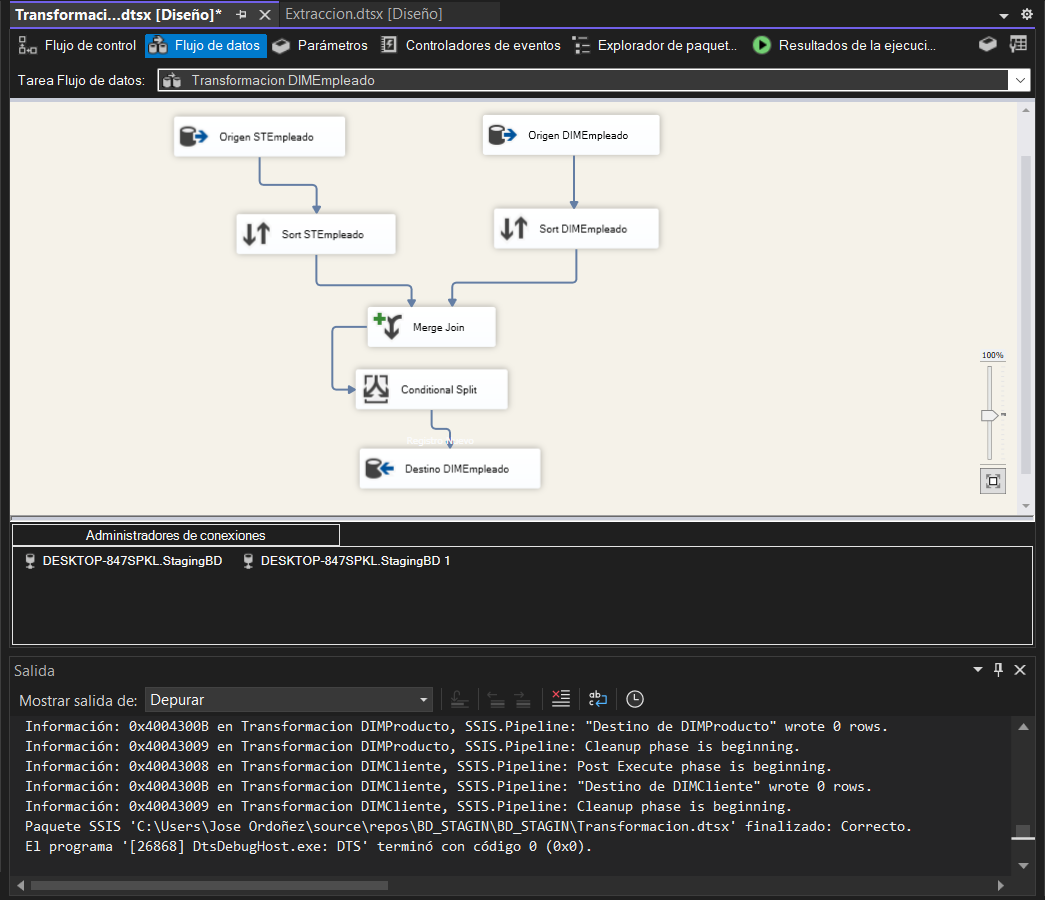
TRANSFORMACION

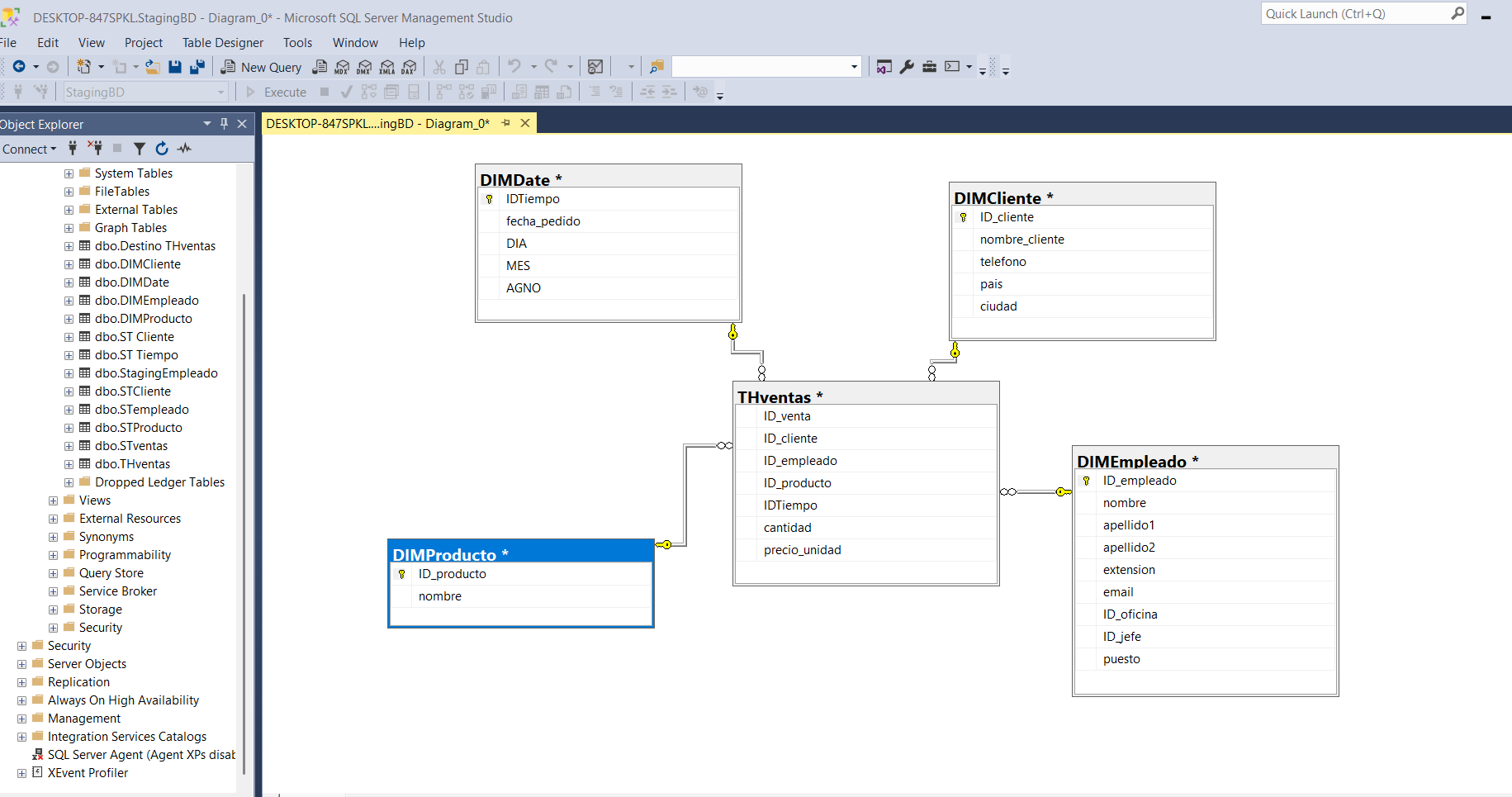


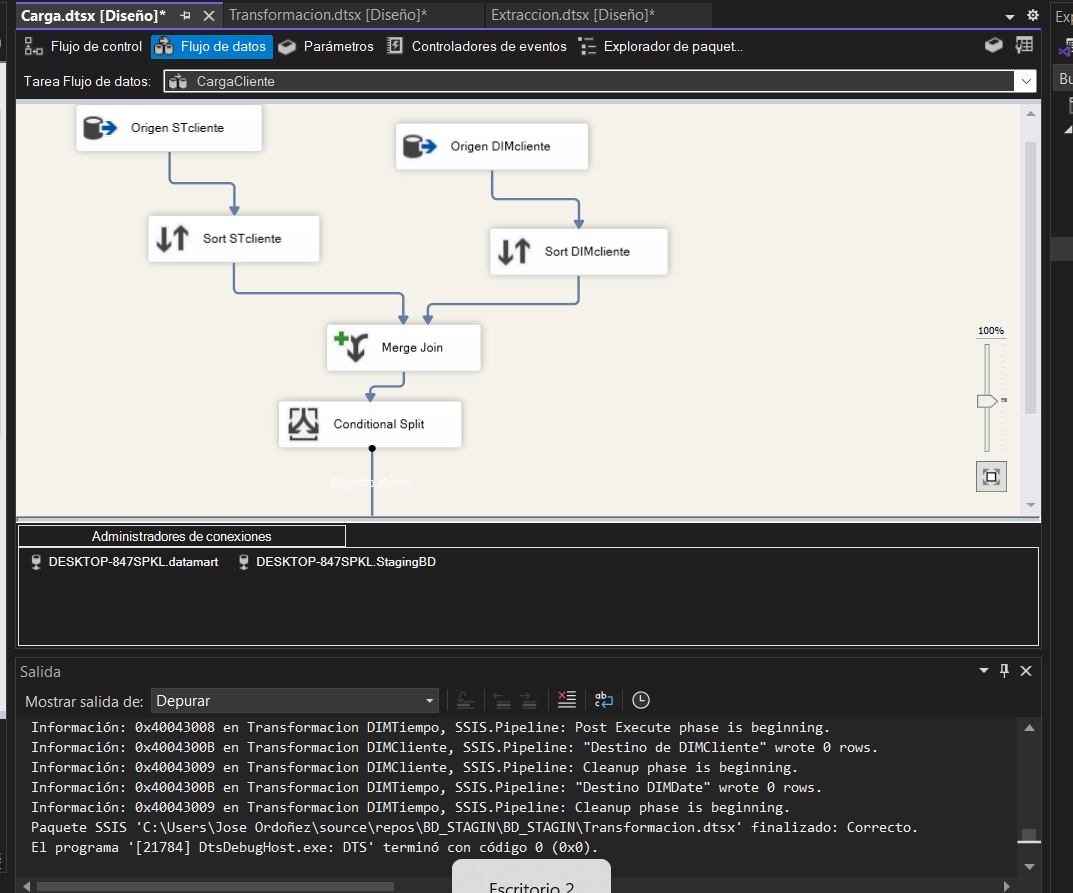


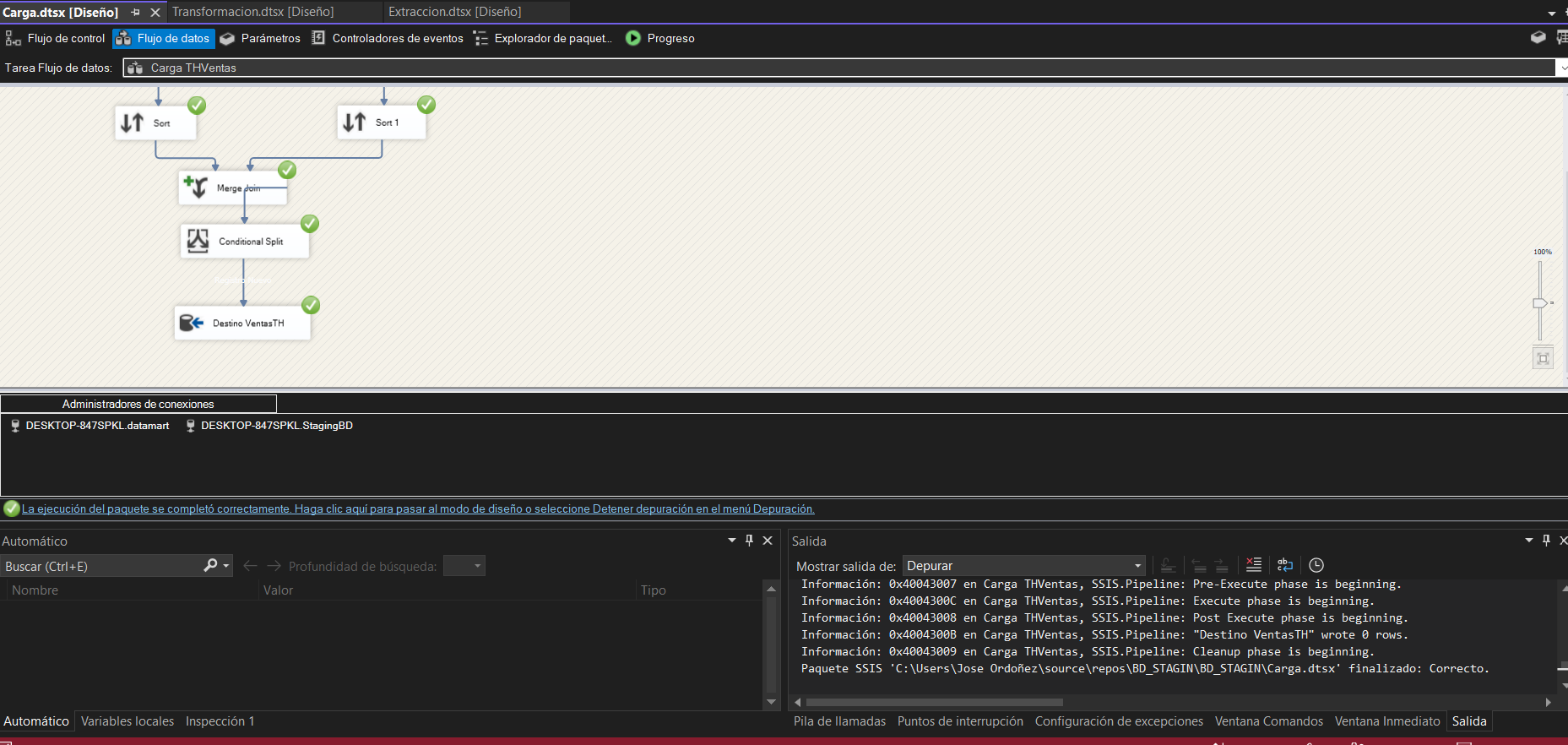












# Conclusiones

La implementación de una base de datos Staging para la BD Jardinería permite mejorar significativamente la calidad y organización de los datos, abordando problemas como la falta de estandarización, la duplicidad de registros y los errores en la captura de información. Gracias a esta solución, se garantiza la integridad y coherencia de los datos, optimizando la generación de reportes y facilitando la toma de decisiones basada en información precisa y confiable.

Además, la incorporación de una capa intermedia de limpieza y validación de datos minimiza las inconsistencias en los reportes gerenciales y mejora la integración con otros sistemas. Esto no solo optimiza el desempeño del sistema de información actual, sino que también sienta las bases para futuras expansiones y mejoras en el manejo de datos.

El proceso de Extracción, Transformación y Carga (ETL) implementado permitió convertir datos operacionales en información estructurada y confiable, lista para el análisis de negocio. A través de técnicas de limpieza, normalización y enriquecimiento, se garantizó la calidad de los datos, eliminando inconsistencias y estandarizando formatos para su correcta explotación en el Data Mart.

# Bibliografía.

<https://pdfs.semanticscholar.org/f08f/aceeb7f74648430357d97ee217fa3d1ba484.pdf>

<https://www.obsbusiness.school/blog/como-utilizar-el-modelo-estrella-en-una-base-de-datos>

<https://bibliotecadigital.usb.edu.co/server/api/core/bitstreams/5e2faccb-f2ca-4595-9701-24a80de15107/content>

<https://aws.amazon.com/es/what-is/data-mart/>