**ACTIVIDAD 2**

# Tema

Proceso de transformación de datos y carga en el data mart

**Integrantes**

Juan Pablo Bedoya

Ferney Rodrigo Marin

**Docente**

Antonio Jesús Valderrama

# Materia

Bases de datos 2

Ingeniería de software y Datos

**2025**

# Descripción de la actividad

**Instrucciones:**

Desarrollo de proceso de transformación de datos y carga en el *Data Mart* final.

1. **Preparación:**
   1. Revisar el modelo Estrella definido en el proyecto de análisis de la base de datos Jardinería para comprender la estructura y las relaciones entre las tablas de dimensiones y la tabla de hechos. [*Consúltala aquí*](https://iudigital.instructure.com/courses/19646/files/6157478/download).
   2. Verificar la disponibilidad y consistencia de la base de datos de *staging* previamente creada para el proyecto.
2. **Extracción de datos desde la base de datos origen hacia la base de datos de *Staging*:**
   1. Utilizar consultas SQL para extraer datos relevantes de la base de datos origen y cargarlos en las tablas correspondientes de la base de datos de *staging.*
   2. Verificar la integridad y consistencia de los datos extraídos para asegurar que cumplan con los requisitos del modelo Estrella.
3. **Transformación de datos según las necesidades analíticas:**
   1. Aplicar técnicas de transformación de datos, como limpieza, normalización y enriquecimiento, para preparar los datos de acuerdo con las necesidades analíticas específicas.
   2. Realizar la transformación de los datos utilizando consultas SQL u herramientas de ETL (*Extract, Transform, Load*) según sea necesario para garantizar la calidad y coherencia de los datos.
4. **Carga de registros en el *Data Mart* final:**
   1. Diseñar consultas SQL o scripts de carga para insertar los registros transformados desde la base de datos de *staging* en las tablas del *data mart*final.
   2. Ejecutar las consultas de carga y verificar que los datos se hayan insertado correctamente en el *data mart* final.
5. **Documentación y presentación:**
   1. Documentar todo el proceso de transformación de datos y carga en un informe detallado que incluya una descripción de las etapas realizadas, las consultas SQL utilizadas.
   2. Presentar el informe de manera clara y concisa, asegurándose de incluir referencias al modelo Estrella y las técnicas de ETL aplicadas.

# Introducción

En la actualidad, las empresas deben gestionar grandes volúmenes de datos de manera eficiente para optimizar sus procesos y mejorar la toma de decisiones. La implementación de un proceso de Extracción, Transformación y Carga (ETL) es fundamental para garantizar la calidad de la información y su correcta integración en un modelo de análisis.

Este documento detalla la aplicación de un proceso ETL en la base de datos de Jardinería, abordando cada una de sus fases: extracción de datos desde la base transaccional, transformación de la información para mejorar su calidad y carga en el Data Mart final. Se describen también las estructuras utilizadas, las consultas empleadas y las optimizaciones realizadas para garantizar un análisis eficiente.

# Objetivos

* Identificar los datos clave en la base de datos Jardinería y evaluar su relevancia para el análisis.
* Diseñar e implementar una base de datos Staging como paso intermedio en el proceso ETL.
* Desarrollar consultas SQL para la extracción, transformación y carga de los datos en el Data Mart.
* Asegurar la integridad, consistencia y calidad de los datos antes de su integración en el modelo dimensional.
* Documentar los procesos y estructuras utilizadas para futuras referencias y mejoras.

# Planteamiento del problema

La base de datos **Jardinería** presenta una alta normalización, lo que dificulta la ejecución de consultas analíticas eficientes. Aunque el modelo estrella propuesto optimiza el acceso a los datos, la extracción directa de la base transaccional genera problemas de rendimiento y dificulta la transformación de la información.

Para resolver este problema, se establece una base de datos **Staging** que permitirá depurar, estructurar y optimizar los datos antes de integrarlos en el modelo dimensional definitivo. Este enfoque reduce la carga sobre el sistema transaccional y mejora la calidad de la información utilizada en el análisis.

**Análisis del problema**

La base de datos Jardinería presenta una estructura altamente normalizada, lo que dificulta la ejecución de consultas analíticas de manera eficiente. La extracción directa desde el sistema transaccional puede generar problemas de rendimiento y dificultar la transformación de la información.

Para abordar esta situación, se implementó un proceso ETL con una base de datos Staging, que permite depurar, estructurar y optimizar los datos antes de integrarlos en el modelo dimensional definitivo. Este enfoque reduce la carga sobre la base transaccional y mejora la calidad de los datos utilizados en el análisis.

## Ajustes realizados en la entrega 1

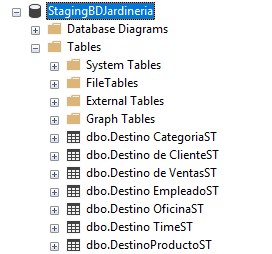
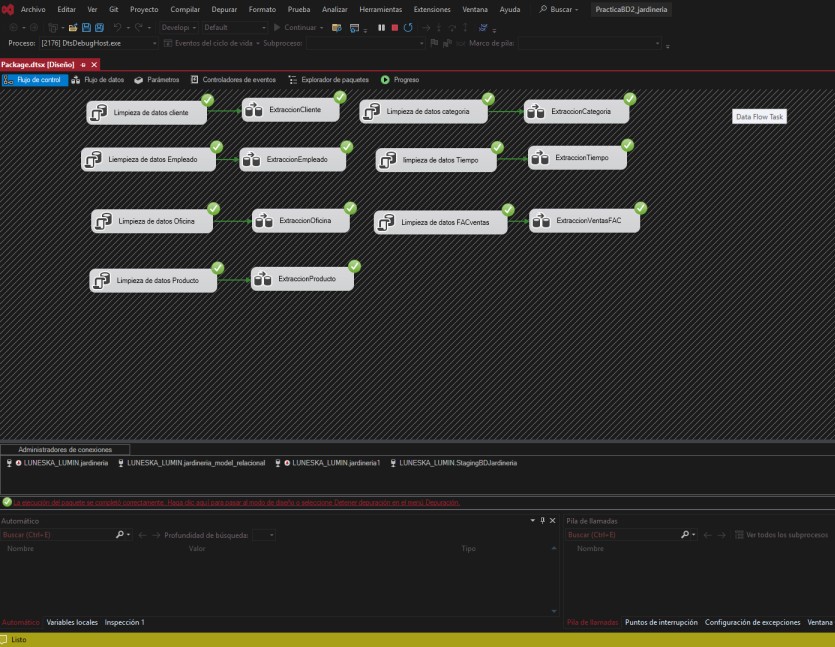
Luego del análisis detallado de **Jardinería** y la implementación de **Staging**, se realizaron las siguientes modificaciones:

* Incorporación de un proceso ETL que permite la limpieza y transformación de datos antes de su integración.
* Refinamiento de la estructura de las dimensiones para evitar redundancias innecesarias.
* Ajuste de claves primarias y foráneas en **Staging** para asegurar integridad referencial.

## Proceso de traslado de datos a Staging

El procedimiento de migración de datos se realizó en tres etapas:

* **Extracción**: Identificación y obtención de los datos relevantes mediante consultas SQL.



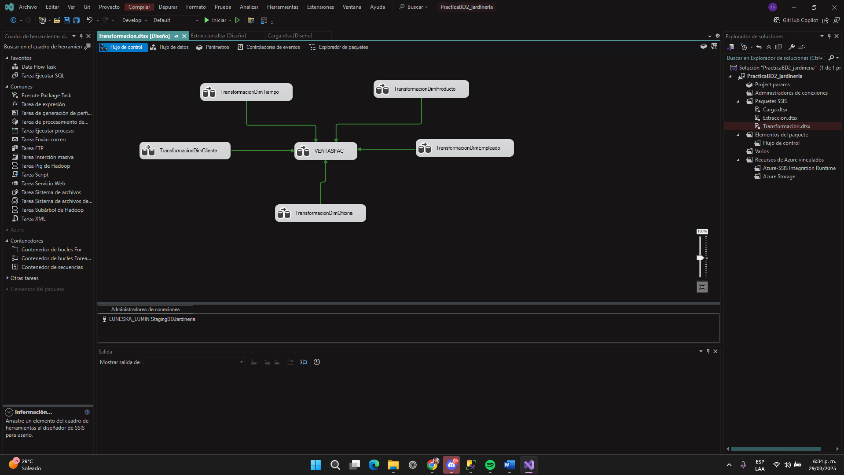
## Estructura de la base de datos Staging

1. **Destino de VentasST** o ID\_pedido (INT) o ID\_producto (INT) o ID\_cliente (INT) o fecha\_pedido (DATE) o ID\_empleado (INT) o ID\_oficina (INT) o cantidad (INT) o precio\_unidad (NUMERIC(15,2)) o descuento (NUMERIC(29,4)) o impuesto (NUMERIC(29,4)) o total (NUMERIC(31,4)) o Fac (NUMERIC(26,2))
2. **DestinoProductoST** o ID\_producto (INT) o nombre (NVARCHAR(70)) o Categoria (INT) o descripcion (NVARCHAR(MAX)) o cantidad\_en\_stock (SMALLINT) o precio\_venta (NUMERIC(15,2))
3. **Destino OficinaST** o ID\_oficina (INT) o Descripcion (NVARCHAR(10))

o ciudad (NVARCHAR(30)) o pais (NVARCHAR(50)) o region (NVARCHAR(50)) o codigo\_postal (NVARCHAR(10)) o telefono (NVARCHAR(20)) o linea\_direccion1 (NVARCHAR(50)) o linea\_direccion2 (NVARCHAR(50))

1. **Destino CategoriaST** o Id\_Categoria (INT) o Desc\_Categoria (NVARCHAR(50)) o descripcion\_texto (NVARCHAR(MAX))
2. **Destino TimeST** o Fecha (NVARCHAR(51))
3. **Destino de ClienteST** o ID\_cliente (INT) o nombre\_cliente (NVARCHAR(50)) o nombre\_contacto (NVARCHAR(30)) o apellido\_contacto (NVARCHAR(30)) o telefono (NVARCHAR(15)) o fax (NVARCHAR(15)) o linea\_direccion1 (NVARCHAR(50)) o linea\_direccion2 (NVARCHAR(50)) o ciudad (NVARCHAR(50)) o pais (NVARCHAR(50))
4. **Destino EmpleadoST** o ID\_empleado (INT) o nombre (NVARCHAR(50)) o apellido1 (NVARCHAR(50)) o apellido2 (NVARCHAR(50)) o extension (NVARCHAR(10)) o email (NVARCHAR(100)) o ID\_oficina (INT) o ID\_jefe (INT) o puesto (NVARCHAR(50))

* **Transformación**: Limpieza, ajuste de formatos y eliminación de inconsistencias en los datos extraídos.



**Captura de pantalla de un celular

El contenido generado por IA puede ser incorrecto.**

* Se aplicaron técnicas de normalización en la creación de dimensiones, como la tabla DIMCliente en el Data Mart (DIMVentas).
* Se realizaron conversiones de tipos de datos y limpieza de valores nulos.
* Se utilizaron relaciones entre las tablas para garantizar la consistencia de los datos.
* Tablas de transformación:
  + **DIMCliente (**
    - *ID\_cliente INT PRIMARY KEY,*
    - *nombre NVARCHAR(50),*
    - *apellido NVARCHAR(50),*
    - *ciudad NVARCHAR(50),*
    - *pais NVARCHAR(50),*
    - *telefono NVARCHAR(20) )*
  + **DIMProducto (**
    - *ID\_producto INT PRIMARY KEY,*
    - *nombre NVARCHAR(70),*
    - *categoria INT,*
    - *descripcion NVARCHAR(MAX),*
    - *cantidad\_en\_stock SMALLINT,*
    - *precio\_venta NUMERIC(15,2) )*
  + **DIMEmpleado (**
    - *ID\_empleado INT PRIMARY KEY,*
    - *nombre NVARCHAR(50),*
    - *apellido1 NVARCHAR(50),*
    - *apellido2 NVARCHAR(50),*
    - *extension NVARCHAR(10),*
    - *email NVARCHAR(100),*
    - *ID\_oficina INT,*
    - *ID\_jefe INT,*
    - *puesto NVARCHAR(50) )*
  + **DIMOficina** *(*
    - *ID\_oficina INT PRIMARY KEY,*
    - *Descripcion NVARCHAR(10),*
    - *ciudad NVARCHAR(30),*
    - *pais NVARCHAR(50),*
    - *region NVARCHAR(50),*
    - *codigo\_postal NVARCHAR(10),*
    - *telefono NVARCHAR(20),*
    - *linea\_direccion1 NVARCHAR(50),*
    - *linea\_direccion2 NVARCHAR(50) )*
* **Carga**: Inserción de los datos en la base **DIMVentas**, garantizando optimización y coherencia.

Tabla

El contenido generado por IA puede ser incorrecto.**Captura de pantalla de un celular

El contenido generado por IA puede ser incorrecto.**

* Se diseñaron consultas SQL para insertar los registros transformados desde la base de datos de staging en las tablas del Data Mart (**DIMVentas**).
* Se ejecutaron consultas de carga asegurando la integridad de los datos y evitando duplicaciones.
* Tablas del Data Mart final:
  + **DIMCliente**
  + **DIMProducto**
  + **DIMEmpleado**
  + **DIMOficina**
  + **DIMVentas**

# Conclusiones

La implementación del proceso ETL ha permitido mejorar la integración, transformación y carga de los datos desde la base de datos transaccional de Jardinería hasta el Data Mart. A través de la fase de extracción, se consolidaron los datos relevantes en staging; mediante la transformación, se mejoró la calidad, consistencia y estructuración de la información; y finalmente, con la carga, se garantizó que los datos estén disponibles para su análisis eficiente. Este enfoque reduce la carga sobre el sistema de origen, optimiza la precisión de la información y facilita la toma de decisiones basada en datos confiables y bien organizados

## Anexos

Estimado profesor en el siguiente enlace de github se encuentran los archivos que estos contienen el script de la [**scriptDBStagingActualizado**](https://github.com/Emjey25/ScriptBDUniversity/blob/main/script_DBStagingActualizado.sql) y **DIMVentas** las consultas están en otro script junto con los script originales que genera dentro visual estudio al momento de nosotros crear las tablas en esa interfaz , y los otros dos archivos son los **Backups** de las dos bases de datos tanto la original como la staging y la dimventas como la de las Dimensiones.

**Bibliografías**

*¿Qué es ETL? - Explicación de extracción, transformación y carga (ETL) - AWS*. (n.d.).

Amazon Web Services, Inc. <https://aws.amazon.com/es/what-is/etl/>

Que es el proceso de extracción y transformación de carga  [https://www.ibm.com/eses/topics/etl](https://www.ibm.com/es-es/topics/etl)