**ACTIVIDAD 2**

# Tema

Creación de una base de datos staging

**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

Redactar un documento escrito que cumpla con normas APA y contenga:

* Introducción
* Objetivos
* Planteamiento del problema
* Análisis del problema • Propuesta de la solución con:
* Correcciones a la entrega 1.
* Descripción del análisis realizado a los datos *Jardinería* y cómo estos se trasladaron a la base de datos *Staging.*
* Anexos: debe agregar los BK de las dos bases de datos y el documento del *script* de las consultas para crear la base de datos *Staging.*
* Bibliografía.

# Introducción

En la actualidad, las empresas enfrentan el desafío de administrar grandes volúmenes de datos de manera eficiente. La correcta estructuración y análisis de la información se ha convertido en una necesidad fundamental para la toma de decisiones estratégicas. Con el fin de mejorar el procesamiento y la calidad de los datos, este documento expone la implementación de una base de datos Staging, la cual servirá como un puente entre la base transaccional Jardinería y el modelo dimensional propuesto.

Aquí se detalla el análisis de los datos originales, el proceso de extracción, transformación y carga (ETL), y las correcciones implementadas para garantizar la calidad de la información. Asimismo, se documenta la estructura de Staging y las consultas utilizadas para trasladar los datos de manera óptima.

# Objetivos

* Identificar los datos clave en la base de datos **Jardinería** y determinar su relevancia en el análisis.
* Diseñar e implementar una base de datos **Staging** que sirva como área de procesamiento intermedio antes de la integración en el modelo dimensional.
* Desarrollar y ejecutar las consultas necesarias para trasladar la información desde

**Jardinería** a **Staging**, garantizando integridad y coherencia.

* Documentar las correcciones realizadas en relación con la entrega anterior y validar la correcta estructuración de los datos.

# 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**

Las principales dificultades al extraer datos desde **Jardinería** son:

* La fragmentación de la información debido a la normalización extrema.
* La presencia de datos redundantes o inconsistentes que requieren limpieza.
* La necesidad de realizar múltiples uniones para obtener información consolidada.

Para abordar estos desafíos, se diseñó una base de datos **Staging** que almacena los datos en un formato optimizado para su integración en el modelo estrella.

## 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.

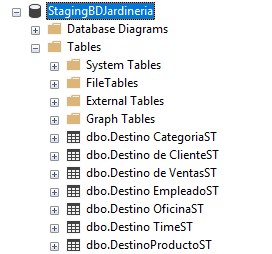
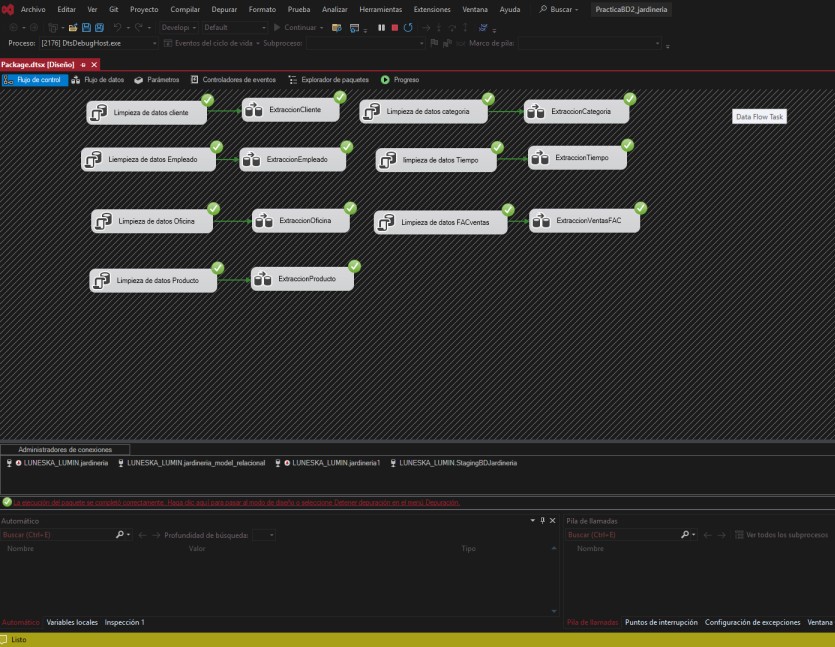
## Proceso de traslado de datos a Staging

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

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

**Cabe resaltar que en este trabajo se evidencia una de las 3 fases, en este trabajo se enfoca en la extracción**.

1. **Transformación**: Limpieza, ajuste de formatos y eliminación de inconsistencias en los datos extraídos.
2. **Carga**: Inserción de los datos en la base **Staging**, garantizando optimización y coherencia.



## 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))

# Conclusiones

La implementación de **Staging** mejora la integración de datos entre **Jardinería** y el modelo estrella, optimizando la calidad y reduciendo la carga sobre el sistema transaccional. Este enfoque permite una mejor estructuración de la información y facilita su uso para análisis posteriores.

## Anexos

Estimado profesor en el siguiente enlace de github se encuentran 4 archivos que estos contienen el script de la bdStaging y 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

<https://github.com/Emjey25/ScriptBDUniversity.git>

**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)