



**IU** Digital  
de Antioquia



## **Evidencia de Aprendizaje 2 : Creación de una base de datos staging**

**Curso : Bases de datos II**

### **Estudiantes**

**Liz Karen Espinosa Cantor**

**Oscar Luis Marquez Arrieta**

**Juan David Ramírez García**

### **Docente:**

**Antonio Jesús Valderrama**

**IU digital de Antioquia**

**Medellín**

**14 de Septiembre del 2025**

## ÍNDICE

Introducción.....	4
Objetivos.....	4
Planteamiento del problema.....	5
Análisis del problema.....	5
Propuesta de la solución.....	6
Descripción del análisis realizado a los datos de Jardinería y cómo estos se trasladaron a la base de datos Staging.....	12
Conclusiones.....	13
Anexos.....	13
Bibliografía.....	13

## Introducción

La idea de este trabajo es aprender qué es una base de datos staging y comprender su importancia dentro del proceso de construcción de un data warehouse. Para esto, utilizaremos la base de datos Jardinería, la cual usamos en la primera actividad de aprendizaje. En este caso, identificamos y analizaremos las tablas más relevantes y campos, con el fin de transformarlas y limpiarlas para llevarlas a la nueva base de datos staging.

La metodología de desarrollo del trabajo corresponde a la propuesta en la descripción de instrucciones de la actividad para la creación de las consultas de una base de datos Staging, las cuales se llevarán a cabo de la siguiente manera:

1. Análisis de la base de datos jardinería para identificar cuáles campos son relevantes para su traslado
2. Construcción de la base de datos Staging, diseño de las tablas, construcción de las consultas y ejecución y validación de las consultas
3. Creación de los Back ups
4. Documentación del proceso de creación

## Objetivos

- **Objetivo general:** Aprender los conceptos y la importancia que conlleva la creación de una base de datos Staging como parte del proceso de un data warehouse. Asimismo, como las buenas prácticas y consideraciones a la hora de hacerlo.
- **Objetivos específicos:**
  1. Identificar las tablas y campos relevantes de la base de datos Jardinería para la construcción de la base de datos Staging.
  2. Analizar la información de la base de datos Jardinería para determinar qué procesos de limpieza y transformación son los más adecuados para el proceso de traslado de los datos.
  3. Diseñar la estructura de la base de datos Staging, definiendo las tablas, sus relaciones y restricciones necesarias para garantizar la integridad y consistencia de los datos.
  4. Ejecutar y validar las consultas desarrolladas, comprobando que los datos se almacenen de forma correcta y completa en la base de datos Staging.
  5. Generar copias de respaldo (Backups) tanto de la base de datos Jardinería como de la base de datos Staging, con el propósito de asegurar la preservación y disponibilidad de la información.
  6. Elaborar la documentación del proyecto bajo normas APA

## Planteamiento del problema

La base de datos Jardinería está conformada por varias tablas no organizadas de una manera óptima, ya que para acceder a cierta información se deben ejecutar consultas haciendo uso de muchos JOINS, es decir, que no permite una consulta de datos de manera directa. Esto debido a que la información se encuentra en relaciones que dificultan su interpretación.

Algunas dificultades encontradas en la base de datos Jardinería cuando se analizó:

- Redundancia de los datos, algunos campos se repiten en distintas tablas como NombreClientes, que aparece tanto en la tabla clientes como en pedidos; esto puede generar duplicidad, ya que un cliente puede estar escrito de diferente forma.
- Campos nulos en la tabla pedidos, el campo FechaEntrega no siempre aparece.
- Dependencia de múltiples tablas para obtener información básica. Es necesario relacionar varias tablas. Por ejemplo, si quisiera saber cuántos productos de una categoría específica han sido pedidos por los clientes, tendría que hacer joins entre las tablas categorías, productos, pedidos y detalles pedidos.

Por lo cual el propósito de hacer la base de datos staging es limpiar, transformar y resumir los datos de manera más eficiente con la intención de trasladar datos mucho más confiables y poder responder a las preguntas de negocio más eficientemente.

## Análisis del problema

La base de datos de Jardinería contiene información de productos, pedidos, clientes y categorías, pero está distribuida en varias tablas con diferentes relaciones entre ellas y no de manera optimizada. Sin embargo, se identifican algunos aspectos que pueden convertirse en retos o problemas a la hora de realizar un proceso de análisis, migración o integración de datos hacia un *Data Warehouse* o una base de datos de *Staging*. Esto dificulta el análisis y entendimiento de las tablas para dar respuesta a preguntas de negocio.

Dificultades encontradas en la base de datos Jardinería:

- La información está dispersa: Para saber cuántas unidades de un producto se vendieron en un año se deben unir mínimo tres tablas como: los nombres de

productos que aparecen en la tabla producto, la cantidad vendida que está en la tabla detalle\_pedido, y la fecha de compra en la tabla pedido.

- Datos nulos: fecha entrega y comentarios en tabla pedido. descripción html y imagen en la tabla Categoría\_producto.
- también se denota que no hay restricciones de unicidad lo que podría generar que un cliente pueda compartir telefono o direccion con otro que podría afectar la carga de los datos staging
- Existen varias llaves Foráneas que generan dependencia entre tablas ejemplo: *empleado → oficina, cliente → empleado, pedido → cliente, detalle\_pedido → producto/pedido*), esto crea problemas para trasladar la información por las jerarquías
- Información no relevante para responder las preguntas de la actividad 1: tabla cliente, empleado, oficina y pago
- Poca estandarización en los parámetros de escritura para la referencia de países ejemplo (usa , EE UU y Estados unidos ) todos pertenecientes a la misma referencia geográfica pero sin estándar crearía varios datos que ensucian la base inicial

Para resolver estos problemas, se construye una base de datos staging, que consolida y organiza la información, dejando los datos listos para análisis posteriores.

### **Propuesta de la solución**

Para resolver los problemas de datos dispersos, nulos y no relevantes de la base de datos jardinería y llevarlos a la base de datos staging fueron los siguientes:

- De la tabla detalle\_pedido se tomaron los campos cantidad y precio\_unidad, para calcular las ventas.
- Para la solución del respeto de jerarquías en las llaves foráneas se planteó un proceso de carga incremental y jerárquica iniciando con las tablas independientes (*oficina, categoria\_producto*) y finalizando con las dependientes (*detalle\_pedido, pago*).
- Para solucionar riesgo de unicidad de los datos se valida la información con emails únicos avalados y teléfonos con formato estándar aplicando la limpieza de la base de datos
- De la tabla pedido se tomó el campo fecha\_pedido para crear la tabla Staging\_tiempo, permitiendo luego identificar el año con más ventas.
- De la tabla producto se seleccionaron ID\_producto, nombre, dimensiones, descripción y categoría, los cuales describen los productos. Estos se relacionan con Desc\_categoria de la tabla categoria\_producto para vincular cada producto a su categoría.
- Para la solución de estandarización geografía se genera un parámetro general donde

se permita unificar los datos y de diferentes referencias enfocadas en una misma región, focalizar a una por ejemplo EEUU, USA se limita a Estados Unidos

- De la tabla categoría\_producto se incluyeron Id\_categoria, Desc\_categoria y descripcion\_texto, los cuales permiten responder a la pregunta de qué categoría tiene más productos.
- A todas las tablas en Staging se les añadió el campo fecha\_carga, con el fin de registrar el momento en que se insertaron los datos, lo que facilita el control y seguimiento del proceso de carga.

A partir de lo anterior las nuevas tablas en Staging Jardinería son:

Staging\_Producto:

- ID\_producto: Tipo entero(int) y es PK el identificador único del producto.
- nombre: Tipo varchar y contiene los nombres de los productos en este caso, productos de jardinería como pala, rastrillo de jardín, etc.
- dimensiones: Tipo varchar y contiene las medidas de cada producto.
- Descripción: Tipo de texto y contiene la descripción de cada producto.
- categoría: Tipo entero que hace referencia a la categoría del producto el cual se describe en la tabla "Dimensión\_categoria"
- Fecha\_carga: tipo date y describe cuándo se realizó la carga a la nueva base de datos, para trazabilidad.

Staging\_Categoria:

- Id\_categoria: Tipo entero (int) y es la clave primaria, el identificador único de la categoría.
- desc\_categoria: Tipo varchar y son los nombres de las categorías como herbáceas y herramientas.
- descripcion\_texto: Tipo de texto y contiene la descripción de cada categoría.
- Fecha\_carga: tipo date y describe cuándo se realizó la carga a la nueva base de datos, para trazabilidad.

Staging\_tiempo

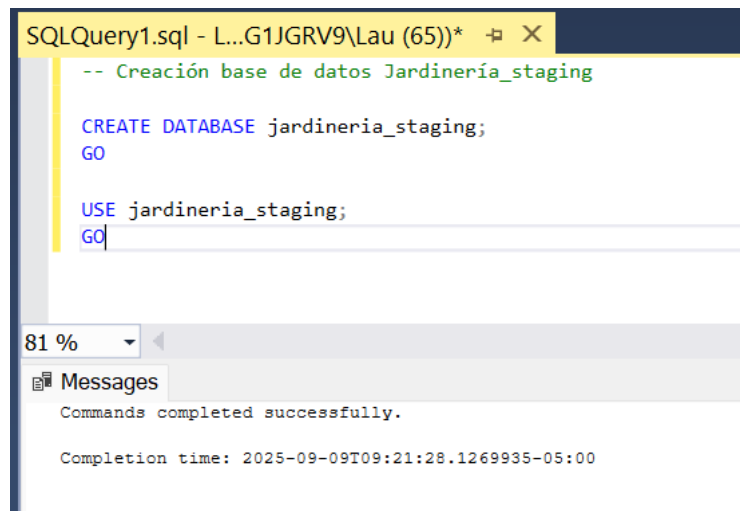
- Id\_tiempo: Es tipo date y es la llave primaria de la tabla, es el identificador único de cada fecha.
- fecha\_pedido: Tipo date es la fecha de pedidos de cada producto y permite calcular periodos de ventas.
- Fecha\_carga: tipo date y describe cuándo se realizó la carga a la nueva base de datos, para trazabilidad.

Staging\_hechos:

- ID\_producto: Tipo entero (int) y es la llave foránea de la tabla Staging\_Producto
- Id\_categoria : Tipo entero (int) y es la llave foránea de la tabla Staging\_Categoria y permite identificar producto por categoría.
- Id\_tiempo: Tipo entero (int) y es la llave foránea con la tabla Staging\_Tiempo y permite identificar pedido de productos por fecha.
- cantidad: Tipo entero(int) y establece el número de los diferentes productos
- precio\_unidad: Tipo float o decimal y contiene el precio de las unidades que se vendieron.
- Fecha\_carga: tipo date y describe cuándo se realizó la carga a la nueva base de datos, para trazabilidad.

**Nota:** Las imágenes corresponden a los scripts usados en Sql Server Management Studio para la construcción y carga de datos en la base de datos Staging.

### Creación de la base de datos Jardinería\_Staging.



The screenshot shows a SQL query window titled 'SQLQuery1.sql - L...G1JGRV9\Lau (65))\*'. The script contains the following commands:

```
-- Creación base de datos Jardinería_staging  
  
CREATE DATABASE jardineria_staging;  
GO  
  
USE jardineria_staging;  
GO
```

Below the script, a 'Messages' pane shows the output: 'Commands completed successfully.' and 'Completion time: 2025-09-09T09:21:28.1269935-05:00'.

### Creación de las tablas con sus campos en la base de datos Jardinería\_staging



```
SQLQuery2.sql - L...G1JGRV9\Lau (66))* X SQLQuery1.sql - L...G1JGRV9\Lau (65))

-- Tabla Staging Producto
CREATE TABLE Producto_staging (
    ID_producto INT PRIMARY KEY,
    nombre NVARCHAR(100),
    dimensiones NVARCHAR(100),
    descripcion NVARCHAR(255),
    categoria INT,
    fecha_carga DATETIME DEFAULT GETDATE()
);

-- Tabla Staging Categoría
CREATE TABLE Categoria_staging (
    Id_categoria INT PRIMARY KEY,
    Desc_categoria NVARCHAR(100),
    descripcion_texto NVARCHAR(255),
    fecha_carga DATETIME DEFAULT GETDATE()
);

-- Tabla Staging Tiempo
CREATE TABLE Tiempo_staging (
    Id_tiempo INT PRIMARY KEY,
    fecha_pedido DATE,
    fecha_carga DATETIME DEFAULT GETDATE()
);

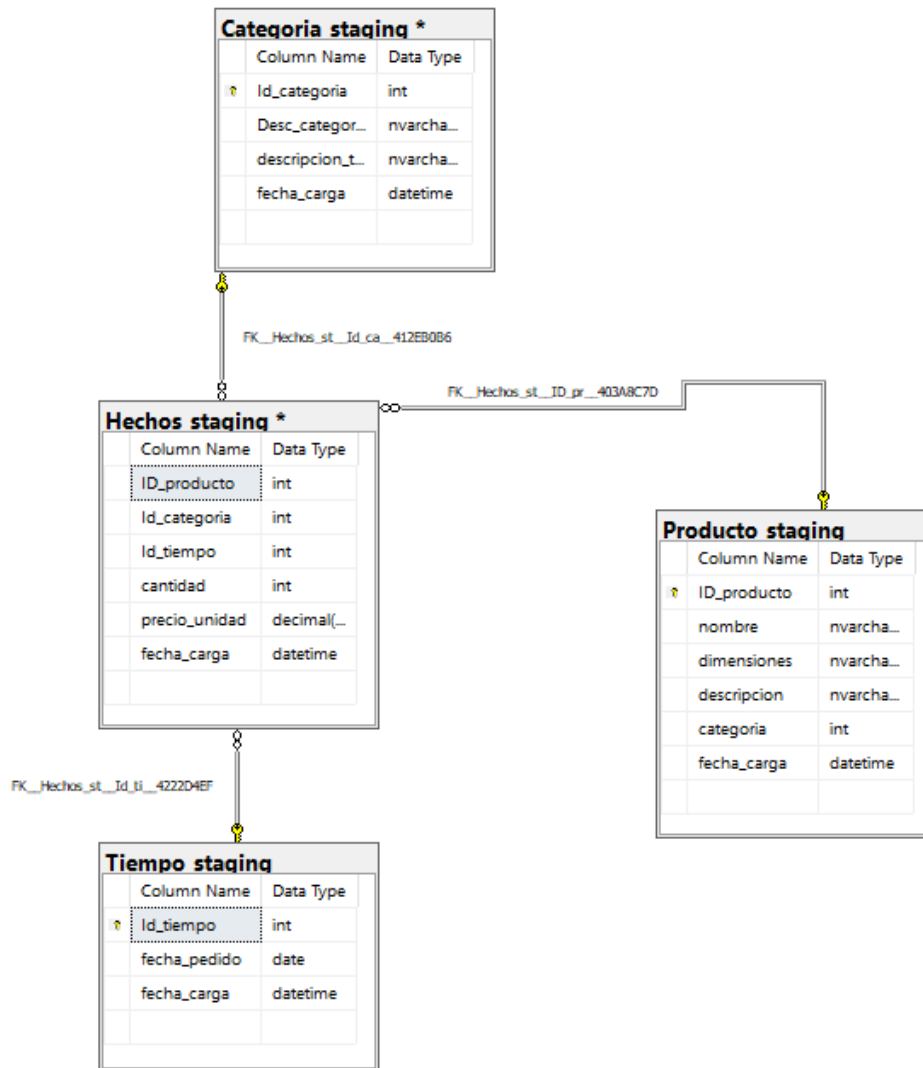
-- Tabla Staging Hechos
CREATE TABLE Hechos_staging (
    ID_producto INT,
    Id_categoria INT,
    Id_tiempo INT,
    cantidad INT,
    precio_unidad DECIMAL(10,2),
    fecha_carga DATETIME DEFAULT GETDATE(),
    FOREIGN KEY (ID_producto) REFERENCES Producto_staging(ID_producto),
    FOREIGN KEY (Id_categoria) REFERENCES Categoria_staging(Id_categoria),
    FOREIGN KEY (Id_tiempo) REFERENCES Tiempo_staging(Id_tiempo)
);

61 %

Messages
Commands completed successfully.

Completion time: 2025-09-09T09:35:28.8253637-05:00
```

Diagrama de relaciones de la base datos Jardinería\_staging



**Proceso de inserción de datos desde la base de datos Jardines hacia la base de datos Staging\_Jardinería:**

Tabla Categoria\_staging

```
LAPTOP-5G1JGRV9...ging - Diagram_0*      SQLQuery2.sql - L...G1JGRV9\Lau (66))* SC
-- Insertar en tabla Categoria_Staging
INSERT INTO Categoria_Staging(Id_categoria, Desc_categoria, descripcion_texto)
SELECT Id_Categoria, Desc_Categoria, descripcion_texto
FROM jardineria.dbo.Categoria_producto;
```

39 %

Messages

(5 rows affected)

Completion time: 2025-09-09T09:58:31.9526962-05:00

SQLQuery3.sql - L...G1JGRV9\Lau (61))*		SQLQuery2.sql - L...G1JGRV9\Lau (66))*	SQLQuer
select * from <u>Categoria_staging</u> ;			
89 %			
Results Messages			
	<u>Id_categoria</u>	<u>Desc_categoria</u>	<u>descripcion_texto</u>
1	1	Herbaceas	Plantas para jardin decorativas
2	2	Herramientas	2 para todo tipo de acción
3	3	Aromaticas	Plantas aromáticas
4	4	Frutales	Árboles pequeños de producción frutal
5	5	Ornamentales	Plantas vistosas para la decoración del jardín
		<u>fecha_carga</u>	2025-09-09 09:58:31.950

Tabla Productos\_staging

SQLQuery2.sql - L...G1JGRV9\Lau (66))*		SQLQuery1.sql - L...G1JGRV9\Lau (65))*	SQLQuery3.sql - L...G1JGRV9\Lau (61))*
<pre>-- Insertar datos en tabla Producto_staging  INSERT INTO <u>Producto_staging</u> (<u>ID_producto</u>, <u>nombre</u>, <u>dimensiones</u>, <u>descripcion</u>, <u>categoria</u>) SELECT ID_producto, nombre, dimensiones, descripcion, Categoria FROM <u>jardineria.dbo.producto</u>;</pre>			
9 %			
Messages			
(276 rows affected)			
Completion time: 2025-09-09T10:25:57.0964767-05:00			
<pre>select * from <u>Producto_staging</u>;</pre>			
89 %			
Results Messages			
	<u>ID_producto</u>	<u>nombre</u>	<u>dimensiones</u>
1	1	Sierra de Poda 400MM	0,258
2	2	Pala	0,156
3	3	Rastrillo de Jardín	1,064
4	4	Azadón	0,168
5	5	Ajedrea	15-20
6	6	Lavándula Dentata	15-20
7	7	Mejorana	15-20
8	8	Melissa	15-20
9	9	Mentha Sativa	15-20
10	10	Petroselinum Hortense (Peregil)	15-20
11	11	Salvia Mix	15-20
12	12	Thymus Citriodora (Tomillo limón)	15-20
13	13	Thymus Vulgaris	15-20
14	14	Santolina Chamaecyparys	15-20
		<u>descripcion</u>	<u>categoria</u>
		Gracias a la poda se consigue manipular un poco la...	2
		Palas de acero con cresta de corte en la punta para...	2
		Fabuloso rastillo que le ayudará a eliminar piedras, ...	2
		Longitud:24cm. Herramienta fabricada en acero y pi...	2
		Planta aromática que fresca se utiliza para condime...	3
		Espliego de jardín, Alhucema rizada, Alhucema dent...	3
		Origanum majorana. No hay que confundirlo con el ...	3
		Es una planta perenne (dura varios años) conocida ...	3
		¿Quién no conoce la Hierbabuena? Se trata de una...	3
		Nombre científico o latino: Petroselinum hortense, P...	3
		La Salvia es un pequeño arbusto que llega hasta el ...	3
		Nombre común o vulgar: Tomillo, Tremoncillo Familia...	3
		Nombre común o vulgar: Tomillo, Tremoncillo Familia...	3
			3
		<u>fecha_carga</u>	2025-09-09 10:25:57.080

Tabla Tiempo\_staging

SQLQuery2.sql - L...G1JGRV9\Lau (66))\*    SQLQuery1.sql - L...G1JGRV9\Lau (65))\*    SQLQuery3.sql - L...G1JGRV9\Lau (61))\*

```
-- Insertar datos en tabla Tiempo_staging

INSERT INTO Tiempo_staging(Id tiempo, fecha_pedido)
SELECT DISTINCT ROW_NUMBER() OVER(ORDER BY fecha_pedido) as Id_tiempo, fecha_pedido from jardineria.dbo.pedido;
```

89 %

Messages

(115 rows affected)

Completion time: 2025-09-09T10:37:01.1942109-05:00

SQLQuery2.sql - L...G1JGRV9\Lau (66))\*    SQLQuery1.sql - L...

```
select * from Tiempo_staging;
```

89 %

Results    Messages

	Id_tiempo	fecha_pedido	fecha_carga
1	1	2006-01-17	2025-09-09 10:37:01.187
2	2	2006-05-25	2025-09-09 10:37:01.187
3	3	2007-01-07	2025-09-09 10:37:01.187
4	4	2007-03-19	2025-09-09 10:37:01.187
5	5	2007-05-20	2025-09-09 10:37:01.187
6	6	2007-06-20	2025-09-09 10:37:01.187
7	7	2007-10-05	2025-09-09 10:37:01.187
8	8	2007-10-23	2025-09-09 10:37:01.187
9	9	2008-01-04	2025-09-09 10:37:01.187
10	10	2008-03-05	2025-09-09 10:37:01.187
11	11	2008-03-10	2025-09-09 10:37:01.187

Tabla Hechos\_staging

-- Insertar datos en tabla Hechos\_staging

```

INSERT INTO Hechos_staging(ID_producto, Id_categoria, Id_tiempo, cantidad, precio_unidad, fecha_carga)
SELECT
    dp.ID_producto,
    p.Categoria AS Id_categoria,
    t.Id_tiempo,
    dp.cantidad,
    dp.precio_unidad,
    GETDATE() AS fecha_carga
FROM jardineria.dbo.detalle_pedido dp
JOIN jardineria.dbo.producto p
    ON dp.ID_producto = p.ID_producto
JOIN jardineria.dbo.pedido pe
    ON dp.ID_pedido = pe.ID_pedido
JOIN jardineria_staging.dbo.Tiempo_staging t
    ON pe.fecha_pedido = t.fecha_pedido;
    
```

89 %

Messages

(629 rows affected)

Completion time: 2025-09-09T10:49:38.1115784-05:00

```
select * from Hechos_staging;
```

89 %

Results Messages

	ID_producto	Id_categoria	Id_tiempo	cantidad	precio_unidad	fecha_carga
1	87	4	1	10	70.00	2025-09-09 10:49:38.087
2	151	5	1	40	4.00	2025-09-09 10:49:38.087
3	165	5	1	25	4.00	2025-09-09 10:49:38.087
4	265	5	1	15	19.00	2025-09-09 10:49:38.087
5	276	5	1	23	14.00	2025-09-09 10:49:38.087
6	57	4	8	3	29.00	2025-09-09 10:49:38.087
7	58	4	8	7	8.00	2025-09-09 10:49:38.087
8	164	5	8	50	4.00	2025-09-09 10:49:38.087
9	165	5	8	20	5.00	2025-09-09 10:49:38.087
10	183	5	8	12	6.00	2025-09-09 10:49:38.087
11	251	5	8	67	64.00	2025-09-09 10:49:38.087
12	271	5	8	5	462.00	2025-09-09 10:49:38.087
13	66	4	15	120	9.00	2025-09-09 10:49:38.087
14	146	5	15	32	5.00	2025-09-09 10:49:38.087
15	147	5	15	11	5.00	2025-09-09 10:49:38.087
16	237	5	15	30	266.00	2025-09-09 10:49:38.087

## Descripción del análisis realizado a los datos de Jardinería y cómo estos se trasladaron a la base de datos Staging.

Este análisis se realizó tomando como referencia la base de datos de jardinería, base de datos principal. Fueron creadas las tablas Categoría Staging, Hecho Staging, Producto Staging y Tiempo Staging, con las cuales se organizó la nueva base de datos staging. En esta nueva base de datos se agregó un nuevo campo llamado **fecha\_carga**; este nos permite guardar la fecha de los datos al momento de entrar al Staging.

Estos datos se trasladaron de la base de datos Jardinería a la de Staging, obteniendo información valiosa;

- De la tabla detalle\_pedido se tomaron los campos cantidad y precio\_unidad, está para calcular las ventas.
- Así mismo se tomó de la tabla pedido el campo fecha\_pedido para crear la tabla tiempo Staging y poder identificar el año con más ventas.
- de la tabla producto se tomo ID\_producto, nombre, dimensiones, descripcion y categoría, los cuales describen los productos, relacionados con Desc\_categoria de la tabla categoria\_producto y así poder vincular cada producto a su categoría.
- La tabla categoría\_producto se incluye Id\_categoria. Desc\_categoria y descripcion\_texto, estos permiten responder a la pregunta de qué categoría tiene más productos.

Con toda la información obtenida de la base de datos de jardinería, se procedió a realizar la base de datos de Staging, se realizó cada una de las tablas de Staging y el nuevo campo de fecha\_carga, se estudió y analizó cada uno de los puntos importantes para el correcto desarrollo y funcionamiento del proyecto. Después se pudo probar agregando respectivamente datos o descripción a cada uno de los campos de cada tabla que se realizó en Staging, tomando como referencia principal la base de datos principal. Todo esto se organizó detalladamente para su correcta construcción.

## Conclusiones

El desarrollo de la base de datos staging nos permitió comprender cómo sería el funcionamiento y la creación de una de estas. De la base de datos anteriormente realizada, se identificaron las tablas más relevantes y se lograron trasladar a la de staging. Esto nos permitió identificar la información importante y poder determinar los procesos de limpieza y transformación para un adecuado traslado de datos.

Al construir una base de datos **Staging**, nos permite evidenciar en un análisis más

profundo factores como **centralizar, limpiar y estandarizar** la información proveniente de Jardinería, resolviendo problemas de duplicidad, campos incompletos y formatos no uniformes sin perder las dependencias jerárquicas de los datos.

Adicional nos permite crear estrategias diferenciadas frente al tratamiento de información no estructurada de manera que no afecten el análisis transaccional de la información.

El crear copias de seguridad o Backups en ejercicios como este nos permite generar una buena práctica de seguridad para proyectos futuros donde se priorice la conservación de los datos y junto con esto, el tener una buena presentación del informe según los parámetros establecidos por el profesor nos brinda formalidad académica, transparencia asegura la trazabilidad del proceso de cada uno, habilidad que nos permitirá crecer en el desarrollo profesional y laboral a futuro.

### Anexos

- Script que contiene el código de cómo se creó la base de datos staging\_Jardineria, las tablas e insertaron los datos. Nombre del archivo: *Script\_creacion\_jardineria\_staging*  
**Nota para su uso:** Para que funcione el Script se debe ejecutar por partes:  
Primero, crear la base de datos jardineria\_staging; segundo, crear las tablas Staging; y tercero, insertar datos en las tablas. (Archivo se encuentra en la carpeta EA2\_GrupoBD2025\_staging\_Jardineria que vino con este documento)
- Backups de la base de datos Jardinería y Staging\_Jardineria (Los archivos se encuentran en la carpeta EA2\_GrupoBD2025\_staging\_Jardineria que vino con este documento)

### Bibliografía

- GeeksforGeeks. (2024, 5 de agosto). What is a Data Staging Area in a data warehouse?. Recuperado el 10 de septiembre de 2025, de <https://www.geeksforgeeks.org/software-testing/what-is-a-data-staging-area-in-data-warehouse/>
- ByteByteGo. (2025, 8 de enero). What is a staging area (6 Importance of Staging Area in a Data Warehouse) [Video]. YouTube. Recuperado el 10 de septiembre de 2025, de [https://www.youtube.com/watch?v=kqi1QT\\_mfbY](https://www.youtube.com/watch?v=kqi1QT_mfbY)