

## **BASES DE DATOS II**

**POR:**

**SEBASTIAN BEDOYA GIRALDO**

**GRUPO:**

**PREICA2401B010093**

**ACTIVIDAD 2**

**EA2: CREACIÓN DE UNA BASE DE DATOS DE STAGING**

**PROFESOR:**

**VICTOR HUGO MERCADO RAMOS**

**INSTITUCIÓN UNIVERSITARIA DIGITAL DE ANTIOQUIA**

**2024**

## **Introducción**

En el mundo empresarial competitivo de hoy, el análisis de datos es fundamental para alcanzar el éxito. Transformar datos sin procesar en información procesable es fundamental para impulsar la toma de decisiones estratégicas, optimizar operaciones y lograr un mayor rendimiento.

Dado que es crucial, reconocemos la importancia de gestionar los datos eficientemente y por ello hemos iniciado el desarrollo de una base de datos para Jardinería. El propósito de esta iniciativa es mejorar la calidad de los datos, optimizar su estructura para el análisis y facilitar la generación oportuna y precisa de informes.

## **Objetivos**

Nuestros objetivos al construir la base de datos de staging son: Nuestros objetivos al construir la base de datos de staging son:

- Mejorar la calidad de los datos: Garantizar la precisión de los datos al eliminar redundancias, inconsistencias y valores faltantes.
- Optimizar la estructura de datos: Mejorar el rendimiento del análisis y la generación de informes mediante la implementación de un modelo de estrella.
- Facilitar el análisis de datos: Posibilitar la extracción eficaz y veloz de información relevante de los datos.
- Mejorar la toma de decisiones: Brindar información exacta y puntual para apoyar la toma de decisiones estratégicas y operativas.

## **Planteamiento del problema**

La analítica de datos es imprescindible para el éxito de cualquier empresa en la actualidad. No obstante, la estructura y organización de las bases de datos existentes a menudo dificultan la capacidad para extraer información significativa de los datos.

La base de datos Jardinería, no es adecuada para un análisis eficiente de datos ya que está diseñada principalmente para operaciones transaccionales. Generar informes y obtener insights accionables se dificulta debido a su estructura rígida, la presencia de redundancias e inconsistencias, y la falta de un modelo de datos orientado al análisis.

## **Necesidades:**

Para superar estos desafíos, se requiere la creación de una base de datos de staging que:

Implemente un modelo de datos orientado al análisis: Utilizar un modelo de estrella para hacer más fácil la agregación de datos, el análisis multidimensional y la generación eficiente de informes

Elimine redundancias e inconsistencias: Asegurar la calidad e integridad de la información mediante la implementación de procesos de limpieza y transformación de datos

Ofrezca flexibilidad para el análisis ad-hoc: Facilitar la generación de ideas inesperadas y respuestas a preguntas emergentes del negocio permitiendo explorar los datos desde varias perspectivas.

Crear una base de datos de staging para Jardinería no consiste en copiar la información actual, sino en modificarla y organizarla para que se convierta en un recurso estratégico para tomar decisiones.

### **Análisis del problema**

Realizamos un análisis en profundidad de la base de datos Jardinería para comprender su estructura, identificar áreas de mejora y determinar los requisitos para la base de datos de staging. Este análisis reveló lo siguiente:

Tablas principales: Las tablas "Producto", "Cliente", "Tiempo", "Pedido" y "Empleado" son esenciales para el análisis, ya que contienen información sobre las transacciones, los clientes, los empleados y los productos, respectivamente.

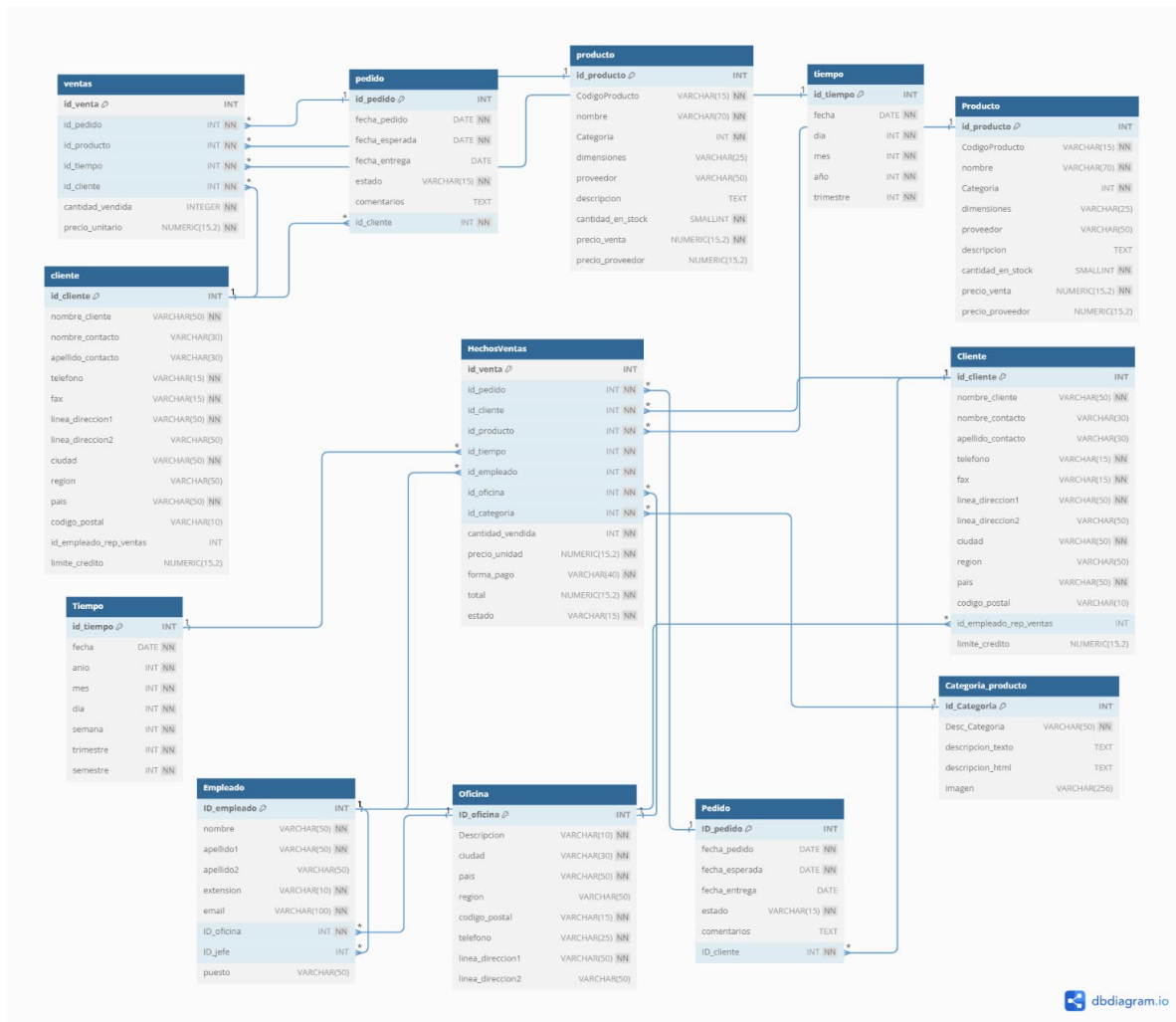
Relaciones entre tablas: Comprender las relaciones entre las tablas, como la relación uno a muchos entre "Pedido" y "Cliente", es crucial para definir las claves externas en la base de datos de staging.

Redundancias e inconsistencias: Se encontraron instancias de datos duplicados e inconsistencias en varias tablas, lo que destaca la necesidad de limpieza y transformación de datos durante el proceso de staging.

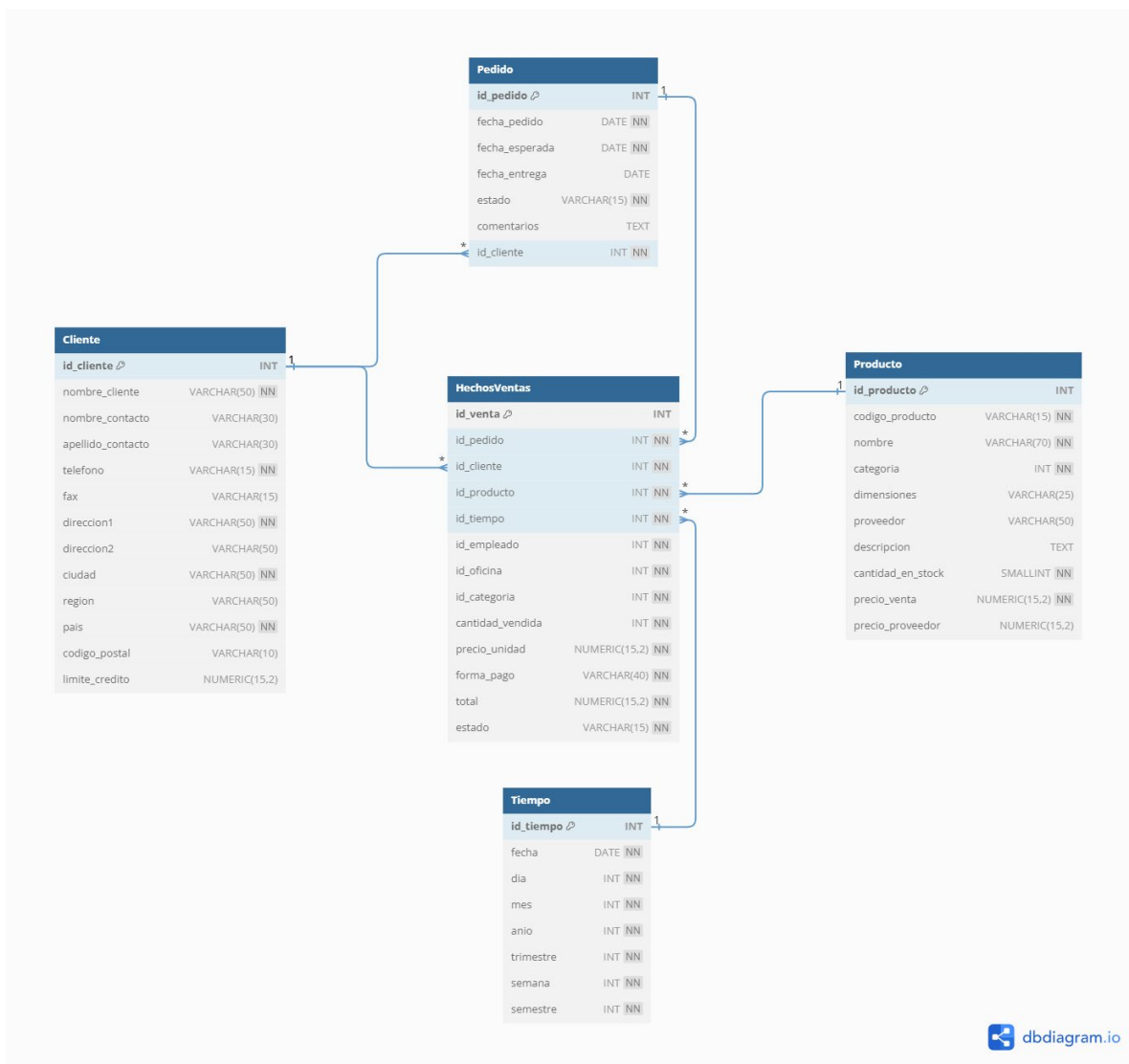
### **Propuesta de solución**

#### **Modelo de estrella:**

### **MODELO ESTRELLA – PRIMERA ENTREGA**



## MODELO ESTRELLA – JARDINERIA CORREGIDA



Se elige un modelo estrella para la base de datos de staging ya que proporciona una estructura optimizada para analizar los datos. La tabla "**HECHOSVENTAS**" es el núcleo de este modelo y contiene las métricas clave del negocio, tales como la cantidad de pedidos, el precio total y el descuento. Esta tabla de hechos está conectada a varias tablas de dimensiones, incluyendo:

Dim\_Cliente: Muestra la fecha del pedido con sus atributos de año, mes y día.

Dim\_producto: Incluye datos de los clientes, como nombre, dirección y detalles para contactar.

Dim\_tiempo: Guarda información de los empleados, incluyendo su nombre, puesto y fecha de inicio.

Dim\_pedido: Proporciona información sobre los productos, incluyendo el nombre, categoría, precio y cantidad disponible.

Dim\_HechosVentas: Captura las transacciones de ventas, actuando como la tabla de hechos central en el modelo de datos estrella.

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

Se llevó a cabo un análisis exhaustivo de cada tabla en la base de datos Jardinería para identificar los campos relevantes para el modelo de estrella.

Antes de cargar los datos en la base de datos de preparación, se realizaron consultas SQL para extraer, limpiar y transformarlos.

Ejemplos de Consultas SQL:

Producto

Campos: id\_producto, nombre, categoria, dimensiones, proveedor, descripcion, cantidad\_en\_stock, precio\_venta, precio\_proveedor.

```
SELECT
  id_producto,
  nombre,
  categoria,
  dimensiones,
  proveedor,
  descripcion,
  cantidad_en_stock,
  precio_venta,
  precio_proveedor
FROM
  Producto
ORDER BY
  id_producto ASC;
```

Cliente

Campos: id\_cliente, nombre\_cliente, nombre\_contacto, apellido\_contacto, telefono, fax, ciudad, region, pais, codigo\_postal, limite\_credito.

```
SELECT
id_cliente,
nombre_cliente,
nombre_contacto,
apellido_contacto,
telefono,
fax,
ciudad,
region,
pais,
codigo_postal,
limite_credito
FROM
Cliente
ORDER BY
id_cliente ASC;
```

## Tiempo

Campos: id\_tiempo, fecha, día, mes, anio, trimestre, semana, semestre.

```
| SELECT
id_tiempo,
fecha,
dia,
mes,
anio,
trimestre,
semana,
semestre
FROM
Tiempo
ORDER BY
fecha ASC;
```

## Pedido

Campos: id\_pedido, fecha\_pedido, fecha\_esperada, fecha\_entrega, estado, comentarios, id\_cliente. También se incluyen detalles de Cliente para relacionar cada pedido con la información del cliente.

```

SELECT
  p.id_pedido,
  p.fecha_pedido,
  p.fecha_esperada,
  p.fecha_entrega,
  p.estado,
  p.comentarios,
  p.id_cliente,
  c.nombre_cliente
FROM
  Pedido p
JOIN
  Cliente c ON p.id_cliente = c.id_cliente
ORDER BY
  p.id_pedido ASC;

```

Además, también se utilizó una tabla de hechos en tus consultas:

## HechosVentas

Campos: id\_venta, cantidad\_vendida, precio\_unidad, forma\_pago, total, estado, además de enlaces a las dimensiones Producto, Cliente, Tiempo, Empleado y Oficina.

```

SELECT
  hv.id_venta,
  hv.cantidad_vendida,
  hv.precio_unidad,
  hv.forma_pago,
  hv.total,
  hv.estado,
  p.nombre AS producto_nombre,
  c.nombre_cliente,
  t.fecha,
  e.nombre AS empleado_nombre,
  o.descripcion AS oficina_descripcion
FROM
  HechosVentas hv
JOIN
  Producto p ON hv.id_producto = p.id_producto
JOIN
  Cliente c ON hv.id_cliente = c.id_cliente
JOIN
  Tiempo t ON hv.id_tiempo = t.id_tiempo
JOIN
  empleado e ON hv.id_empleado = e.id_empleado
JOIN
  oficina o ON hv.id_oficina = o.id_oficina

```



- 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.

Object Explorer

Connect

- LocalDB
  - (localdb)\jardinera SQL Server 15.0.41
    - Databases
      - System Databases
      - Database Snapshots
      - Datos\_Staging
        - Database Diagrams
        - Tables
          - System Tables
          - FileTables
          - External Tables
          - Graph Tables
          - dbo.Categoria\_producto
          - dbo\_cliente
          - dbo\_detalle\_pedido
          - dbo\_empleado
          - dbo.HechosVentas
          - dbo\_oficina
          - dbo\_pago
          - dbo\_pedido
          - dbo\_producto
          - dbo\_tiempo
          - Views
            - External Resources
            - Synonyms
            - Programmability
            - Service Broker
            - Storage
            - Security
              - jardinera

SQLQuery1.sql - (L:\4SCR6T\SEBAS (70))\*

SQLQuery1.sql - (L:\4SCR6T\SEBAS (80))\*

100%

Results Messages

id_producto	nombre	categoria	dimensiones	proveedor	descripcion	cantidad_en_stock	precio_venta	precio_proveedor
1	Sierra de Poda 400MM	2	0.258	HiperGarden Tools	Gracias a la poda se consigue manipular un poco ...	15	14.00	11.00

id_cliente	nombre_cliente	nombre_contacto	apellido_contacto	telefono	fax	ciudad	region	pais	codigo_postal	limite_credito
1		Daniel G	GoldFish	5556901745	5556901746	San Francisco	NULL	USA	24006	3000.00
2		Gardening Associates	Anne	5557410345	5557410346	Miami	NULL	USA	24010	6000.00
3		Garden Valley	Lark	5552321129	5552321128	New York	NULL	USA	25495	12000.00
4		Tendo Garden	Alane	55591233210	55591233211	Miami	NULL	USA	69069	60000.00
5		Lases S.A	Antonio	34916540145	34914851312	Fuenlabrada	Madrid	Spain	28945	154310.00
6		Beragua	Jose	654987321	916549872	Madrid	Madrid	Spain	28942	20000.00
7		Club Golf Puerta del hierro	Paco	62456810	919535678	Madrid	Madrid	Spain	28930	40000.00
8		Naturagua	Gullemo	689234750	916428956	Madrid	Madrid	Spain	28947	32000.00

id_tiempo	fecha	dia	mes	año	tiemstre	semana	semestre
1	2021-01-15	15	1	2021	1	3	1
2	2021-04-10	10	4	2021	2	2	15
3	2021-07-05	5	7	2021	3	27	2
4	2021-10-25	25	10	2021	4	43	2

id_pedido	fecha_pedido	fecha_esperada	fecha_entrega	estado	comentarios	id_cliente	nombre_cliente
16	2009-01-18	2009-02-12	NULL	Pendiente	entregar en murcia	9	DaraDistribuc...
17	2009-01-20	2009-02-15	NULL	Pendiente	NULL	9	DaraDistribuc...
18	2009-01-09	2009-01-09	2009-01-09	Rechaz.	mal pago	9	DaraDistribuc...
19	2009-01-11	2009-01-11	2009-01-13	Entregado	NULL	9	DaraDistribuc...
20	2008-12-30	2009-01-10	NULL	Rechaz.	El pedido fue anulado por el cliente	5	Lases S.A.
21	2008-07-14	2008-07-31	2008-07-25	Entregado	NULL	14	Jardin de Flor...
22	2009-02-02	2009-02-08	NULL	Rechaz.	El cliente carece de saldo en la cue...	1	GoldFish Gar...
23	2009-02-06	2009-02-12	NULL	Rechaz.	El cliente anula la operacion para a...	3	Gerudo Valley

id_venta	cantidad_venta	precio_unidad	foma_pago	total	estado	producto_nombre	nombre_cliente	fecha	empleado_nombre	oficina_descripcion
----------	----------------	---------------	-----------	-------	--------	-----------------	----------------	-------	-----------------	---------------------

## Anexos:

<https://dbdiagram.io/d/6625d53d03593b6b6194902a>

## Conclusión

La capacidad de la empresa para analizar datos y obtener información estratégica se mejora significativamente con la construcción de una base de datos de staging para Jardinería. Durante el proceso de staging, la implementación del modelo estrella, la limpieza de datos y su transformación sientan las bases sólidas para una toma de decisiones informada, optimización operativa y crecimiento empresarial.

## Bibliografía

Kimball, R., & Ross, M. (2013). The data warehouse toolkit: The definitive guide to dimensional modeling (3rd ed.). John Wiley & Sons. Este libro es una referencia fundamental para el modelado dimensional, que se utiliza a menudo en las bases de datos de staging.

Loshin, D. (2011). *Business intelligence: The savvy manager's guide* (2nd ed.). Morgan Kaufmann. Cubre los conceptos básicos de Business Intelligence, incluyendo el papel de las bases de datos de staging.

Han, J., Pei, J., & Kamber, M. (2011). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.