

EVIDENCIA DE APRENDIZAJE 2. CREACIÓN DE UNA BASE DE DATOS DE STAGING

OSCAR DAVID MADRIGAL FONNEGRA
BASES DE DATOS II -IUD

INTRODUCCIÓN

La gestión eficiente de los datos es crucial para mejorar la toma de decisiones, optimizar los procesos comerciales y satisfacer las necesidades de los clientes. Para lograrlo, es fundamental contar con una infraestructura de bases de datos bien diseñada que permita almacenar, procesar y analizar la información de manera efectiva. En este proyecto, nos enfocamos en mejorar la gestión de datos de la bd Jardinería, mediante la construcción de una base de datos Staging que facilite el proceso de Extracción, Transformación y Carga (ETL) de los datos provenientes de su base de datos principal.

OBJETIVOS

OBJETIVO GENERAL:

- Diseñar e implementar una base de datos Staging que permita la extracción de datos desde la base de datos principal de Jardinería, con el fin de facilitar el análisis y la toma de decisiones.

OBJETIVOS ESPECÍFICOS:

1. Realizar un análisis de los datos almacenados en la base de datos de Jardinería para identificar la información relevante y necesaria para el proceso de ETL.
2. Diseñar la estructura de las tablas que conformarán la base de datos Staging, teniendo en cuenta los requisitos de almacenamiento y las relaciones entre los datos.
3. Construir consultas SQL para extraer y cargar los datos desde la base de datos de Jardinería hacia la base de datos Staging.

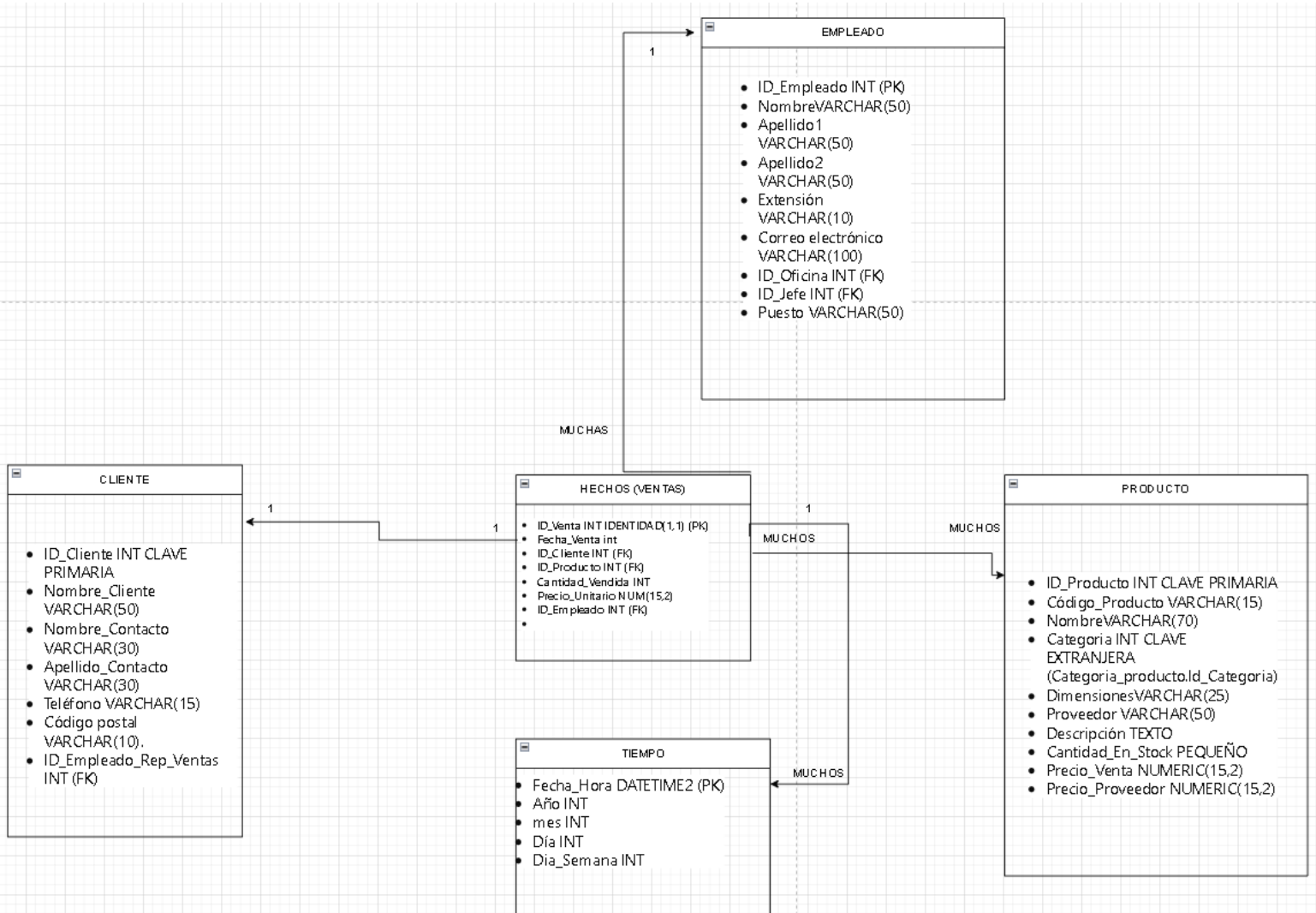
PLANTEAMIENTO DEL PROBLEMA

La gestión de los datos en lo que respecta a la extracción, transformación y carga desde base de datos principal hacia entornos de análisis y reporting hace necesario crear una estructura adecuada para este proceso y así llevar a cabo los análisis precisos y oportunos, lo que puede afectar la toma de decisiones y la eficiencia operativa de la empresa. Por lo tanto, es necesario diseñar e implementar una solución que permita optimizar el manejo de los datos y mejorar la capacidad de análisis de la empresa.

ANÁLISIS DEL PROBLEMA

El análisis de los datos almacenados en la base de datos de Jardinería revela la necesidad de una solución que facilite la extracción y transformación de estos datos para su posterior análisis. Se identifica la falta de una estructura centralizada y estandarizada para el almacenamiento de datos y se implementa en un formato de dimensiones, así como la ausencia de un proceso formalizado para la extracción de datos hacia entornos de análisis. Esto dificulta la integración de datos de diferentes fuentes, la realización de análisis históricos y la generación de informes precisos y oportunos. Además, se observa la presencia de datos redundantes y desactualizados, lo que afecta la integridad y la calidad de la información.

1. Análisis de los datos almacenados en la base de datos Jardinería con sus correcciones anteriores :
 - Revisar los datos almacenados en el Jardinería para identificar cuáles son relevantes y cuáles se deben trasladar a la base de datos *Staging*.



2. Construcción de la base de Datos *Staging*:

- Diseñar la estructura tablas que estarán en la base de datos *Staging*.

Primero creamos la tabla de Hechos en en la base de datos origen en sql server

```
CREATE TABLE Ventas (  
    ID_Venta INT IDENTITY(1,1) PRIMARY KEY,  
    Fecha_Venta DATE,  
    ID_Cliente INT,  
    ID_Producto INT,  
    Cantidad_Vendida INT,  
    Precio_Unitario NUMERIC(15, 2),  
    ID_Empleado INT,  
  
    );
```

y creamos nuestra tabla tiempo

```
CREATE TABLE Tiempo (  
    Fecha DATE PRIMARY KEY,  
    Anio INT,  
    Mes INT,  
    Dia INT,  
    Dia_Semana INT  
);
```

Luego de esto vamos a crear nuestra staging StaginJardineria, la cual vamos a utilizar para nuestra ETL en visual studio

Las consultas para dicho proceso son las siguientes donde realizaremos y trasladaremos los datos más relevantes

Tabla clientes:

```
SELECT id_cliente,nombre_cliente, nombre_contacto, apellido_contacto, telefono,  
linea_direccion1 as Direccion,  codigo_postal, limite_credito  
  
FROM cliente;
```

Tabla producto:

```
SELECT ID_producto, nombre, Categoria, dimensiones, proveedor, descripcion,  
cantidad_en_stock, precio_venta, precio_proveedor  
  
FROM producto;
```

Tabla Ventas:

```
select  
ID_venta, Fecha_Venta, ID_Cliente, ID_Producto, Cantidad_Vendida, Precio_Unitario, ID_Empleado
```

from Ventas

Tabla Tiempo:

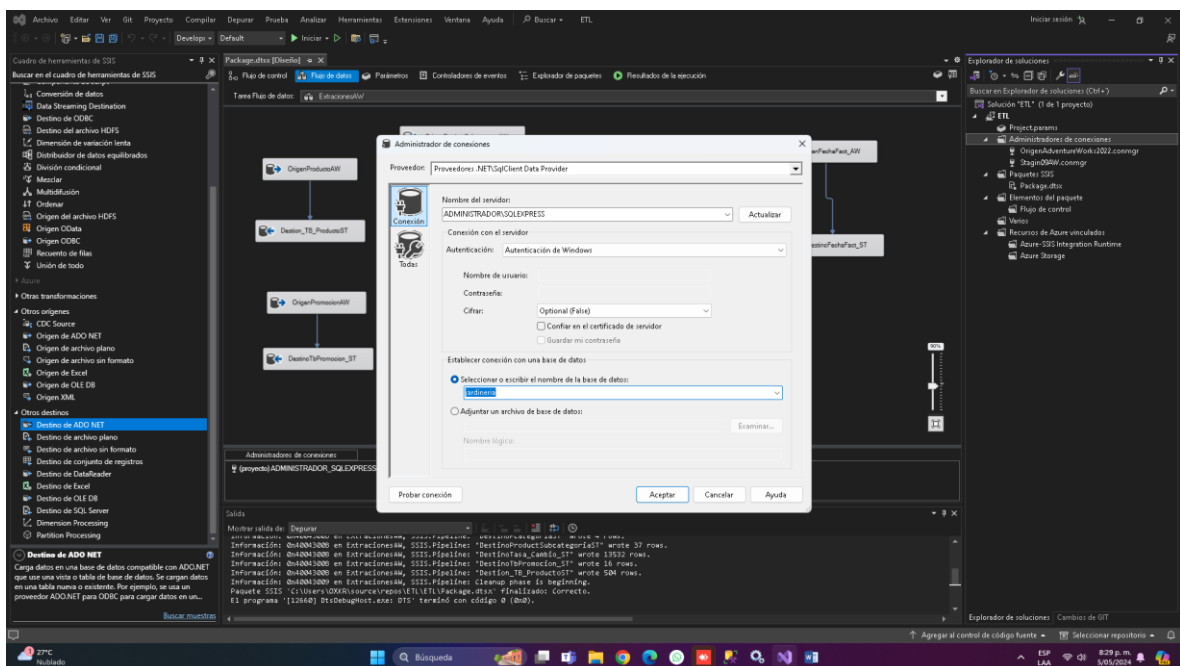
```
select id_tiempo, Anio, Mes, Dia, Dia_Semana  
from tiempo
```

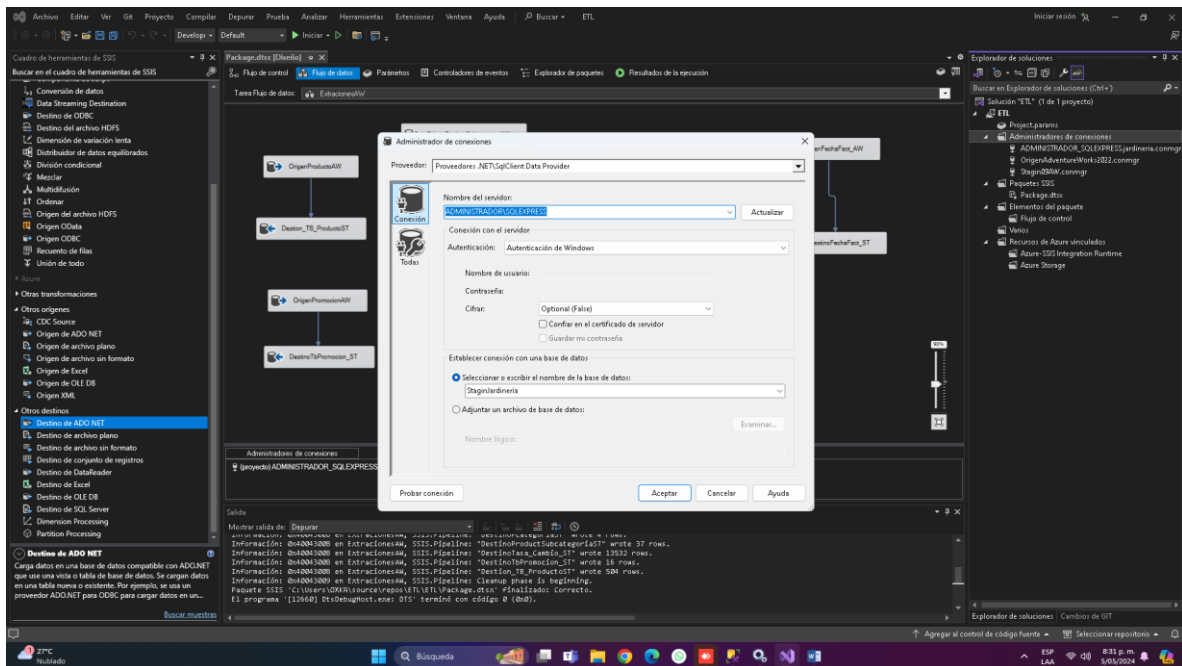
Tabla Empleado:

```
SELECT id_empleado, nombre, apellido1, apellido2, ID_oficina, ID_jefe, puesto  
FROM empleado;
```

3. Ahora creamos las conexiones en el Visual Studio para la base Jardinería –Origen

Y para el destino StaginJardineria





4. Ahora vamos a creamos el Flujo de Control y el Flujo de datos para cada una de las Tablas.

Vamos hacer un ejemplo solo por uno del origen al destino, y así se hace con cada una de las sentencias Sql de consultas de cada tabla, luego mostraremos cuando ejecutemos el Flujo completo en el Visual

```
CREATE TABLE "DestinoVentas_ST" (

    "ID_venta" int identity(1,1),

    "ID_venta" int_0,

    "Fecha_Venta" date,

    "ID_Cliente" int,

    "ID_Producto" int,

    "Cantidad_Vendida" int,

    "Precio_Unitario" numeric(15,2),

    "ID_Empleado" int,

    Primary Key( "ID_venta")

)
```


CREATE TABLE "DestinoEmpleado_ST" (

"id_empleado" int identity(1,1),

"id_empleado_0" int,

"nombre" nvarchar(50),

"apellido1" nvarchar(50),

"apellido2" nvarchar(50),

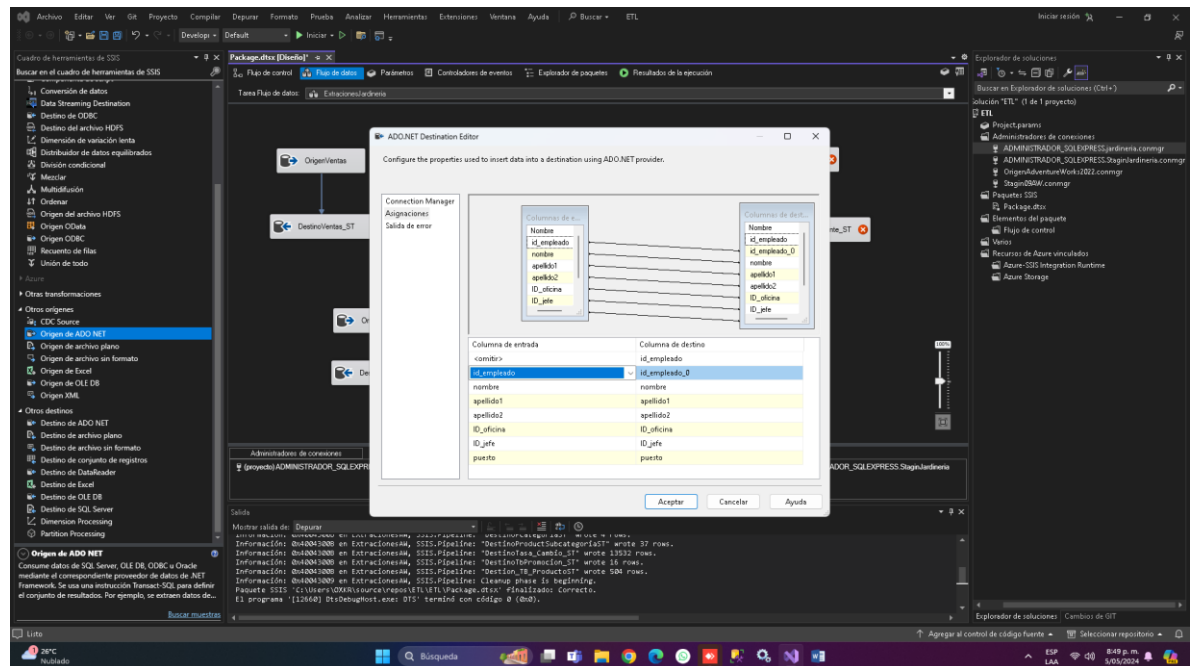
"ID_oficina" int,

"ID_jefe" int,

"puesto" nvarchar(50)

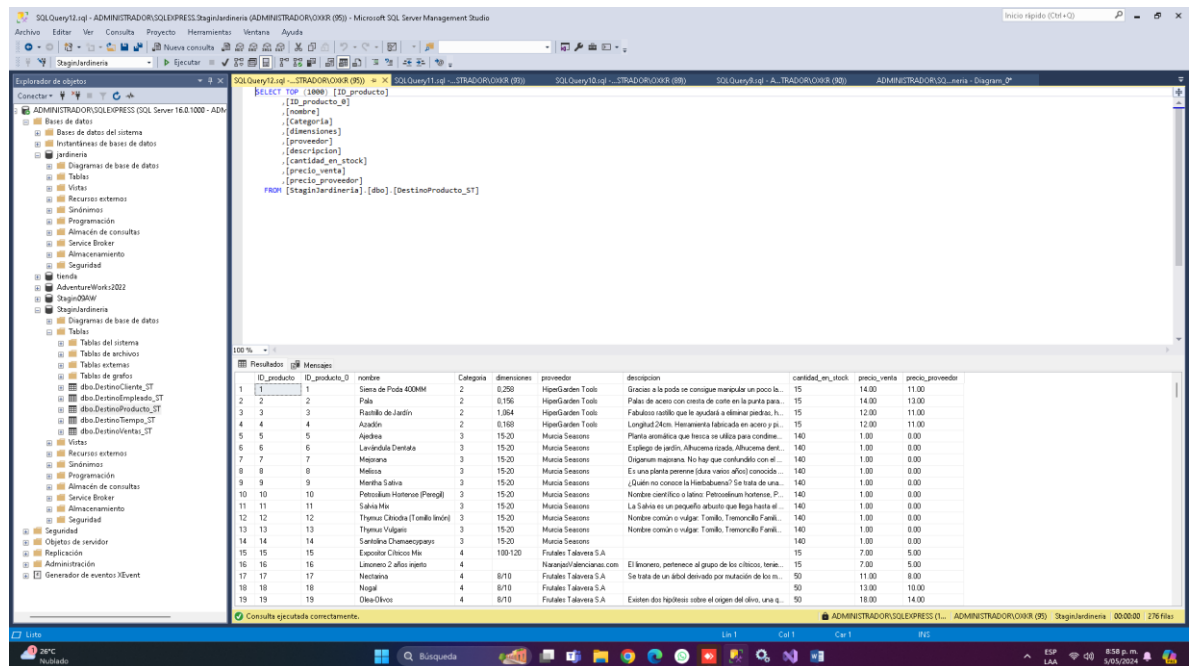
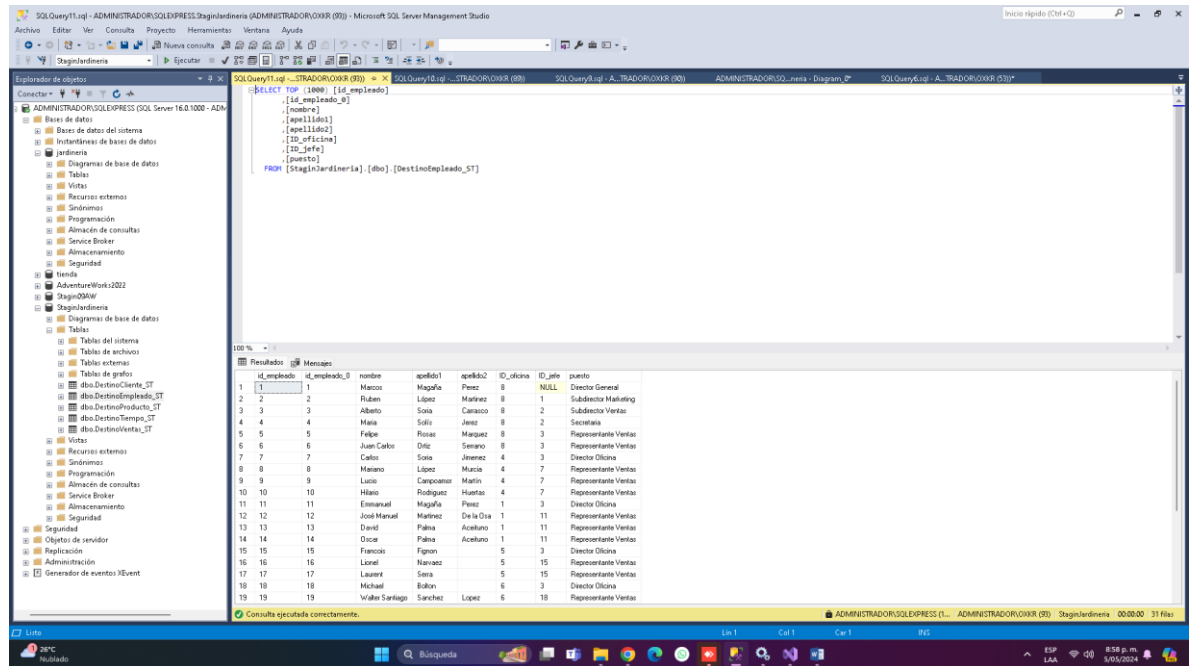
PRIMARY KEY("id_empleado")

)



The screenshot displays the Microsoft SQL Server Enterprise interface. The top window shows the 'Query Editor' with a SELECT statement filtering records from the 'dbo.Stagindadineria' table based on 'id_cliente'. Below the editor, the 'Results' pane shows the execution plan, which includes a 'Table Scan' operation. The bottom pane displays the query results as a table with columns: id_cliente, nombre_cliente, nombre_contacto, apellido_contacto, telefono, direccion, codigo_postal, and limite_credito.

	id_cliente	nombre_cliente	nombre_contacto	apellido_contacto	telefono	Direccion	codigo_postal	limite_credito
1	1	GolfFish	Daniel G	GolfFish	9556007485	Falsa Street 52 A	24006	3000.00
2	2	Gardening Associates	Anne	Wright	5557410345	Wall Avenue	24510	6000.00
3	3	Gawda Valley	Lisa	Raza	5552229129	Dale Avenue #22	85455	1200.00
4	4	Tendo Garden	Akane	Tendo	55591232010	Nail Street #63	656569	60000.00
5	5	Lucas S.A.	Antonio	Lucas	3491540405	C/Lugones 15	28945	154310.00
6	6	Besquis	José	Besquis	654907321	Cipriote square	25942	2000.00
7	7	Club Golf Punta del Bierro	Paco	Lopez	62456810	C/avenida delegada	20350	4000.00
8	8	Naturagus	Guillermo	Rengifo	686224750	C/maedafonda	29947	32000.00
9	9	DraoDistribuciones	David	Serrano	675993000	C/Casas	20946	50000.00
10	10	Madrillob de seguro	José	Tacño	65596545	C/Lugones	29943	20000.00
11	11	Luca S.A.	Antonio	Lucas	3491540405	C/Lugones 15	28945	154310.00
12	12	Canunas Jardines S.L.	Pablo	Canunas	34914872041	C/Vigenes 45	28145	14401.00
13	13	Dardenia S.A.	Juan	Rodriguez	34914553217	C/Rivera York 74	20003	321000.00
14	14	Jardin de Flores	Jaime	Vilar	6545685643	C/Ola 24	20950	4000.00
15	15	Flores Maria	Maria	Rodriguez	660555444	C/Lugones 24	20945	1500.00
16	16	Flores S.A.	Beatriz	Fernandez	606754159	C/Luz Salquibá 4	24566	3500.00
17	17	Naturagarden	Victoria	Cruz	612426529	Piazza Magallan 15	20811	5050.00
18	18	Golf S.A.	Luis	Hernandez	551456762	C/Estrada	26027	30000.00
19	19	Amench Golf Management SL	Mario	Suenz	964453072	C/Laredo	12200	2000.00



Como ven cada una de las tablas quedaron en la nueva base de datos StagingJardineria y la que tienen datos pasaron sus datos sin ningún problema

- Construir el BK de ambas bases de datos.

En el siguiente drive dejaremos los links de las BK de las bases de datos y también el archivo sql de las consultas, además también estarán disponibles en los datos de la carga de la tarea cada uno por separado

https://drive.google.com/drive/folders/1WGpQyeLamvntHOg1_smpRkS_jTj6OcEI?usp=sharing

Bibliografía

- Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. John Wiley & Sons.
- Inmon, W. H., & Hackathorn, R. D. (1994). Using the Data Warehouse. John Wiley & Sons.
- Ralph Kimball et al., (2008). The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. John Wiley & Sons.
- García-Molina, H., Ullman, J. D., & Widom, J. (2008). Database Systems: The Complete Book. Pearson Education.