

Institución Universitaria Digital de Antioquia IUDIGITAL
Faculta de Ingeniería y Ciencias Agropecuarias
Ingeniería de Software y datos

Evidencia de aprendizaje 3. Proceso de transformación de datos y carga en el data
mart final

Omar Aleiser Alvarez Laverde
PREICA2401B010094

Bucaramanga, Santander
26 Mayo 2024

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

Preparación:

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.

- Verificar la disponibilidad y consistencia de la base de datos de staging previamente creada para el proyecto.

Extracción de datos desde la base de datos origen hacia la base de datos de Staging:

- 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.
- Verificar la integridad y consistencia de los datos extraídos para asegurar que cumplan con los requisitos del modelo Estrella.

Transformación de datos según las necesidades analíticas:

- Aplicar técnicas de transformación de datos, como limpieza, normalización y enriquecimiento, para preparar los datos de acuerdo con las necesidades analíticas específicas.
- 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.

Carga de registros en el Data Mart final:

- 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.
- Ejecutar las consultas de carga y verificar que los datos se hayan insertado correctamente en el data mart final.

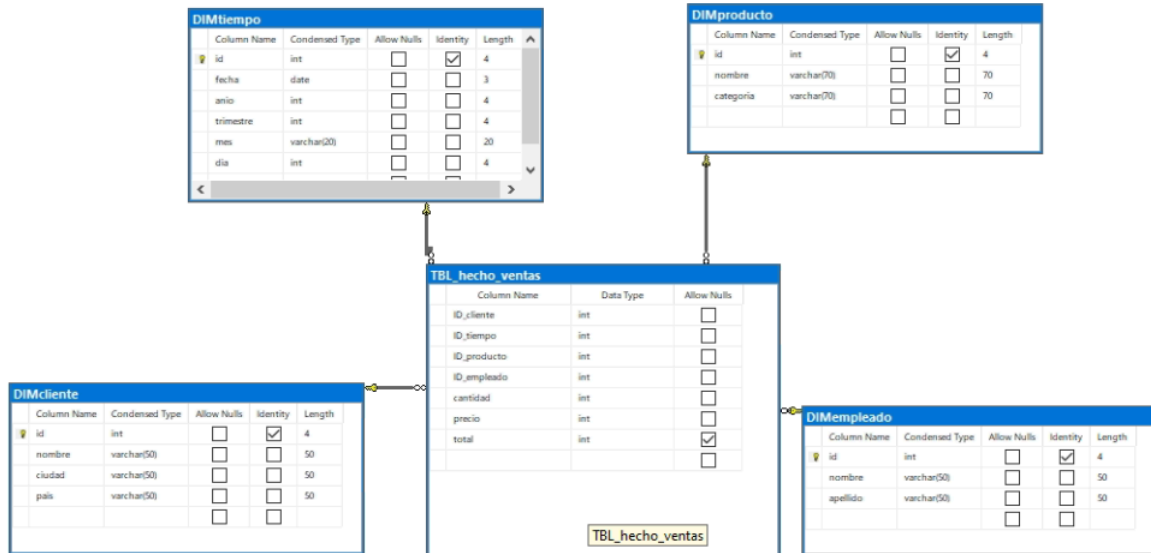
Documentación y presentación:

- 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.
- Presentar el informe de manera clara y concisa, asegurándose de incluir referencias al modelo Estrella y las técnicas de ETL aplicadas.

Desarrollo proyecto

1. Revisar el modelo Estrella

El modelo Estrella generalmente consiste en una tabla de hechos central conectada a varias tablas de dimensiones.



2. Verificar la base de datos de staging

Tabla de hechos: TBL_hechos_ventas

```
CREATE TABLE [dbo].[TBL_hecho_ventas](
    [ID_cliente] [int] NOT NULL,
    [ID_tiempo] [int] NOT NULL,
    [ID_producto] [int] NOT NULL,
    [ID_empleado] [int] NOT NULL,
    [cantidad] [int] NOT NULL,
    [precio] [int] NOT NULL
) ON [PRIMARY]
```

Tabla de dimensiones: Productos, cliente, empleados, tiempo.

```
CREATE TABLE [dbo].[DIMcliente](
    [id] [int] IDENTITY(1,1) NOT NULL,
    [nombre] [varchar](50) NOT NULL,
    [ciudad] [varchar](50) NOT NULL,
    [pais] [varchar](50) NOT NULL,
    CONSTRAINT [PK_DIMcliente] PRIMARY KEY CLUSTERED
```

```
CREATE TABLE [dbo].[DIMempleado](
```

```

[id] [int] IDENTITY(1,1) NOT NULL,
[nombre] [varchar](50) NOT NULL,
[apellido] [varchar](50) NOT NULL,
CONSTRAINT [PK_DIMEmpleado] PRIMARY KEY CLUSTERED

```

```

CREATE TABLE [dbo].[DIMproducto](
    [id] [int] IDENTITY(1,1) NOT NULL,
    [nombre] [varchar](50) NOT NULL,
    [categoria] [varchar](50) NOT NULL,
    CONSTRAINT [PK_DIMproducto] PRIMARY KEY CLUSTERED

```

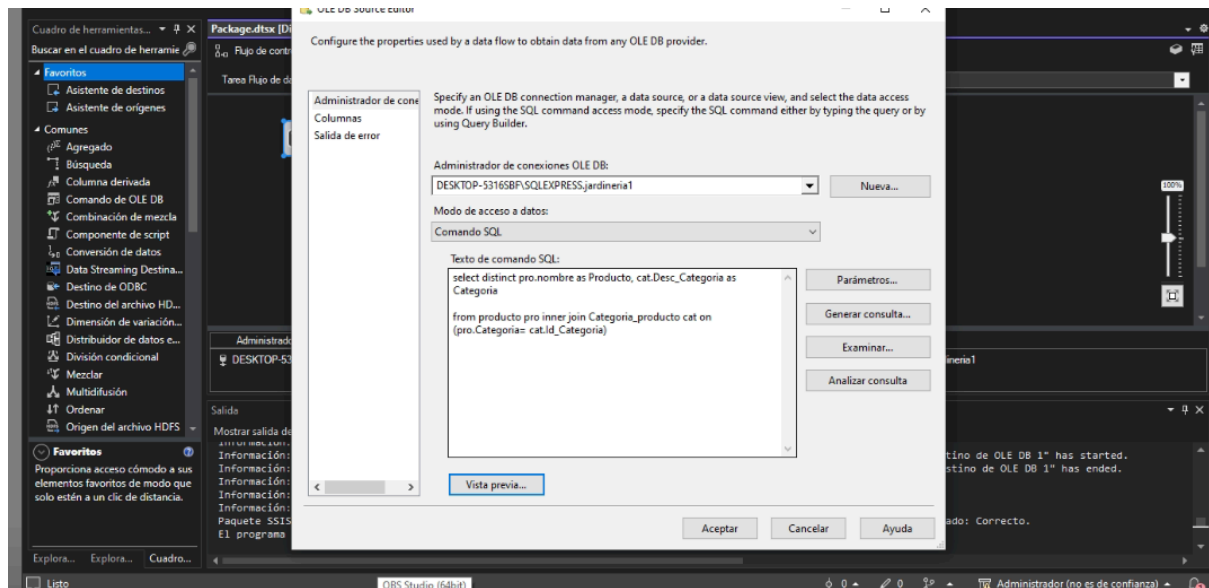
```

CREATE TABLE [dbo].[DIMtiempo](
    [id] [int] NOT NULL,
    [fecha] [date] NOT NULL,
    [anio] [int] NOT NULL,
    [trimestre] [int] NOT NULL,
    [mes] [varchar](20) NOT NULL,
    [dia] [int] NOT NULL,
    CONSTRAINT [PK_DIMtiempo] PRIMARY KEY CLUSTERED

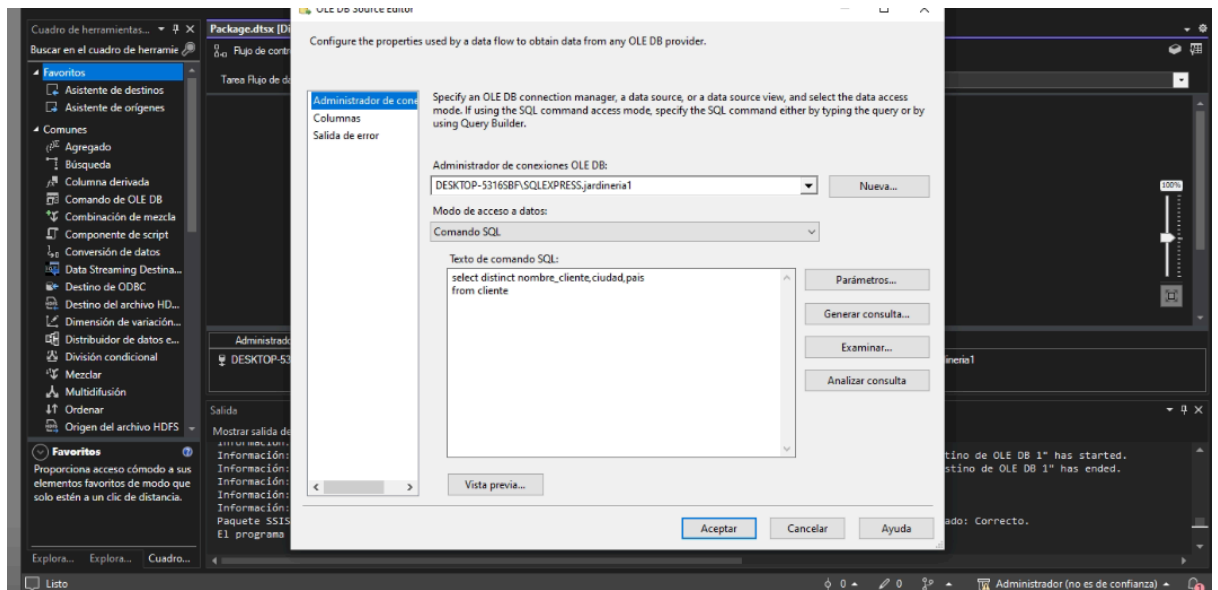
```

Extracción de datos

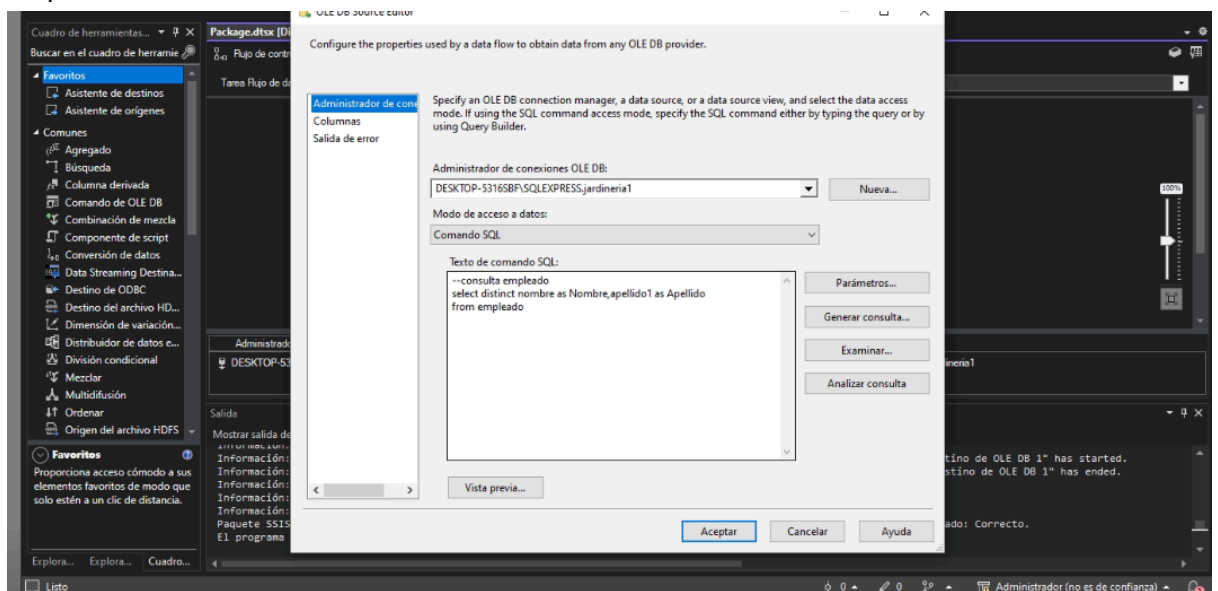
1. Productos



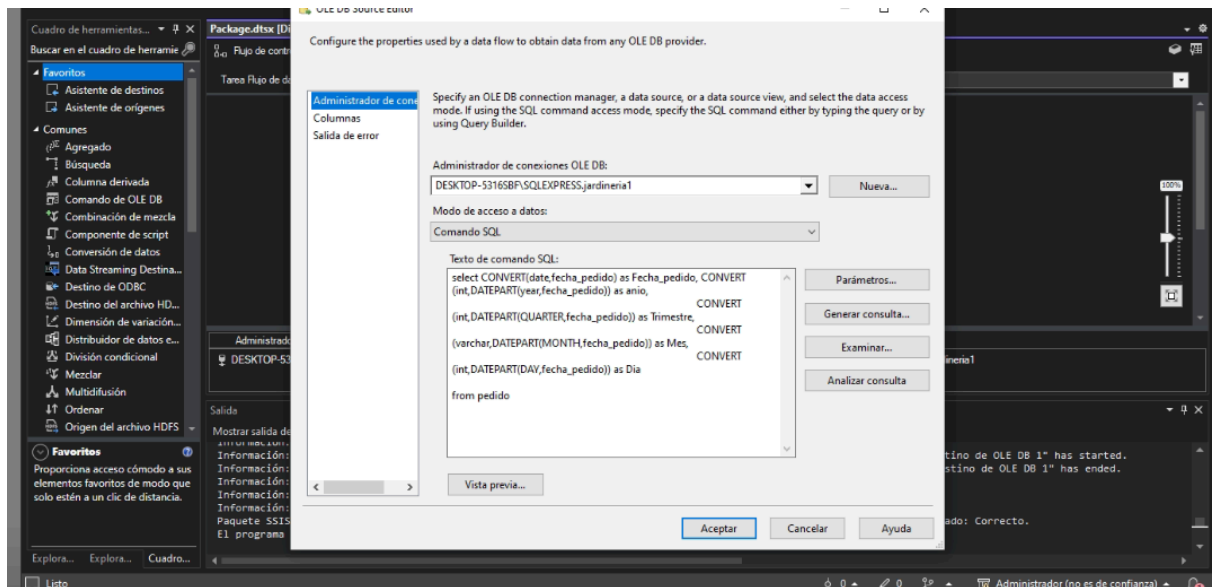
2. Cliente



3. Empleado



4. Tiempo



Verificación de la integridad y consistencia

--Verificar la consistencia de datos

USE datamar_jardineria;

SELECT nombre, categoria, COUNT(*) AS cantidad

FROM DIMproducto

GROUP BY nombre, categoria

HAVING COUNT(*) > 1;

--normalizar los nombres de los productos

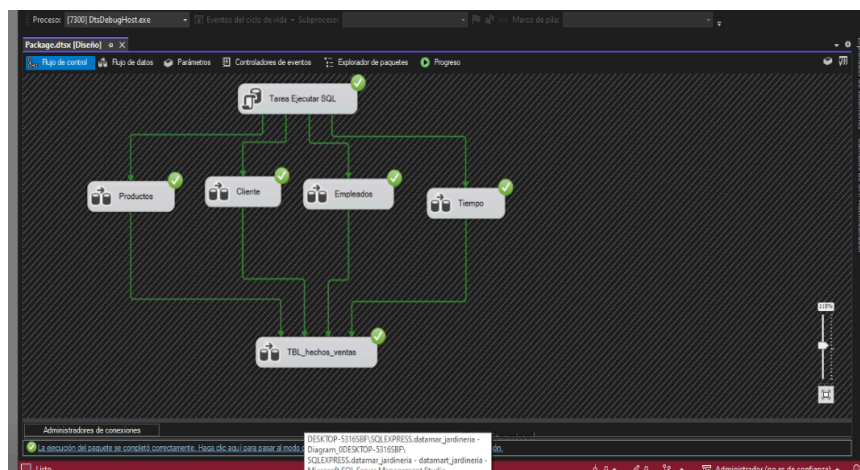
USE datamar_jardineria;

UPDATE DIMproducto

SET nombre = TRIM(UPPER(nombre))

WHERE nombre IS NOT NULL;

Carga de registros en el Data Mart final



Consultas usadas para cargar información al datamart

--Consultar nombre producto y categoría asignada

```
select distinct pro.nombre as Producto, cat.Desc_Categoria as Categoria
```

```
from producto pro inner join Categoria_producto cat on (pro.Categoria= cat.Id_Categoria)
```

-- consultar cliente

```
select distinct nombre_cliente,ciudad,pais  
from cliente
```

--consulta empleado

```
select distinct nombre as Nombre,apellido1 as Apellido  
from empleado
```

--Extraer de la fecha,fecha año, trimestre,mes,día

```
select CONVERT(date,fecha_pedido) as Fecha_pedido,  
CONVERT(int,DATEPART(year,fecha_pedido)) as anio,  
CONVERT(int,DATEPART(QUARTER,fecha_pedido))  
as Trimestre,
```

```
CONVERT(varchar,DATEPART(MONTH,fecha_pedido)) as Mes,  
CONVERT(int,DATEPART(DAY,fecha_pedido)) as Dia
```

```
from pedido
```

--Llenar la data tabla de hechos

SELECT

```
dimC.id AS ID_cliente,  
dimT.id AS ID_tiempo,  
dimP.id AS ID_producto,  
dimE.id AS ID_empleado,  
dp.cantidad,  
dp.precio_unidad,  
dp.cantidad * dp.precio_unidad AS Total
```

FROM jardineria.dbo.detalle_pedido dp

INNER JOIN jardineria.dbo.pedido p ON dp.ID_pedido = p.ID_pedido

INNER JOIN jardineria.dbo.producto prod ON dp.ID_producto = prod.ID_producto

INNER JOIN jardineria.dbo.cliente c ON p.ID_cliente = c.ID_cliente

INNER JOIN jardineria.dbo.empleado e ON c.ID_empleado_rep_ventas = e.ID_empleado

-- Dimensiones

INNER JOIN datamar_jardineria.dbo.DIMcliente dimC ON c.ID_cliente = dimC.id

INNER JOIN datamar_jardineria.dbo.DIMproducto dimP ON prod.ID_producto = dimP.id

INNER JOIN datamar_jardineria.dbo.DIMempleado dimE ON e.ID_empleado = dimE.id

INNER JOIN datamar_jardineria.dbo.DIMtiempo dimT ON p.fecha_pedido = dimT.fecha

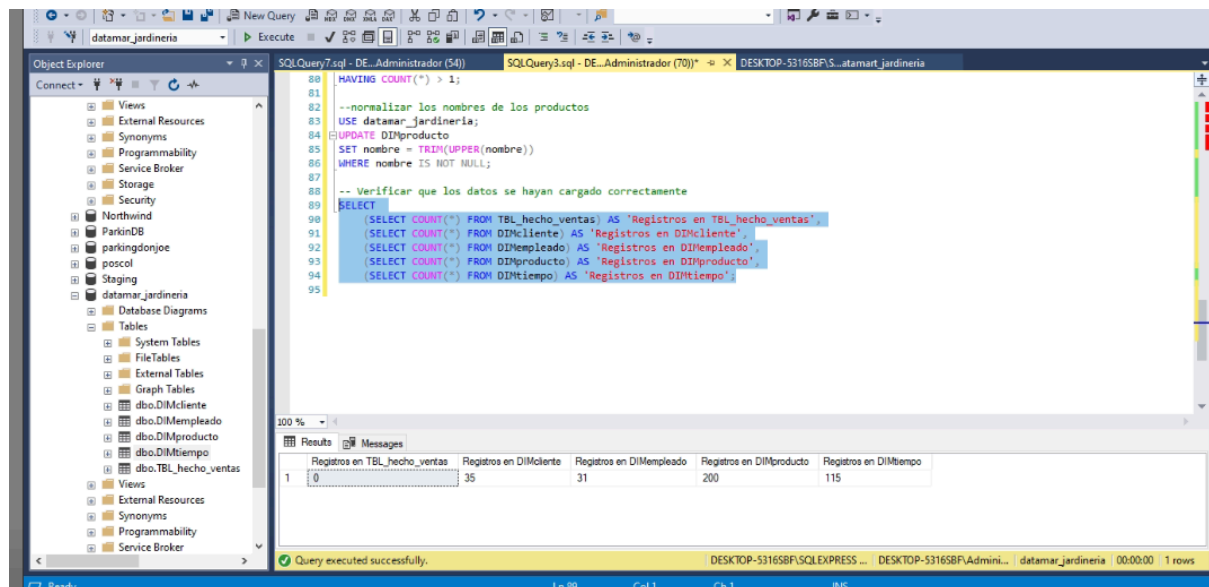
--Eliminar datos del datamart

```
delete from DIMcliente
```

```
delete from DIMempleado
```

delete from DIMproducto
delete from DIMtiempo
delete from TBL_hecho_ventas

Verificar que los datos este cargados correctamente en la base de datos datamart



Script para la consulta si esta la información en la base de datos.

-- Verificar que los datos se hayan cargado correctamente

SELECT

```
(SELECT COUNT(*) FROM TBL_hecho_ventas) AS 'Registros en TBL_hecho_ventas',  
(SELECT COUNT(*) FROM DIMcliente) AS 'Registros en DIMcliente',  
(SELECT COUNT(*) FROM DIMempleado) AS 'Registros en DIMempleado',  
(SELECT COUNT(*) FROM DIMproducto) AS 'Registros en DIMproducto',  
(SELECT COUNT(*) FROM DIMtiempo) AS 'Registros en DIMtiempo';
```

Conclusiones

El proceso ETL se completó exitosamente, y los datos transformados se cargaron en el Data Mart final. Este proceso asegura que los datos estén listos para el análisis posterior.

Link acceso al documento

https://docs.google.com/document/d/190_K5wiaFITilvgvzOTAS0AR66B7m3_-zs3UJdN9dyY/edit?usp=sharing