

Integrantes:

Andres Mauricio Plazas González - PREICA2401B010093
Andrés Rojo Mesa - PREICA2401B010093
Wilmar Alejandro Castañeda Franco - PREICA2401B010076

Programa Académico:
Ingeniería De Software y Datos

Actividades:
S30 - EA3 Proceso de transformación de datos y carga en el data mart final

Asignatura:
Bases de Datos II

Docente:
Víctor Hugo Mercado

Universidad Digital de Antioquia
25 de mayo del 2024

Instrucciones:

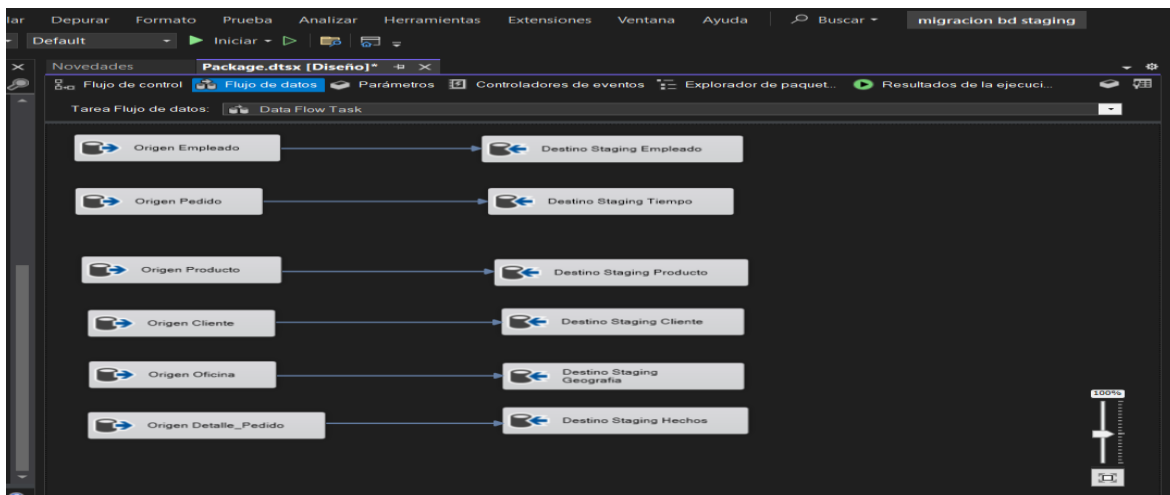
Desarrollo de proceso de transformación de datos y carga en el *Data Mart* final.

1. 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.

2. 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*.



```
USE [jardineria]

//Empleados
SELECT ID_empleado, nombre, apellido1, ID_oficina AS ID_geografia FROM [dbo].[empleado]

//Pedido
SELECT ID_pedido AS ID_tiempo, fecha_pedido, fecha_esperada, fecha_entrega FROM [dbo].[pedido]

//Producto
SELECT ID_producto, CodigoProducto, nombre, Categoria, proveedor,cantidad_en_stock,precio_venta, precio_proveedor FROM [dbo].[producto]

//Cliente
SELECT ID_cliente, nombre_cliente, ID_empleado_rep_ventas, limite_credito FROM [dbo].[cliente]

//geografia
SELECT ID_oficina AS ID_geografia, ciudad, pais, region FROM [dbo].[oficina]

//hechos
SELECT A.ID_detalle_pedido AS ID_venta,
A.cantidad, A.precio_unidad,
A.cantidad*A.precio_unidad AS total,
B.ID_cliente, A.ID_producto, D.ID_empleado_rep_ventas AS ID_empleado,
E.ID_oficina AS ID_geografia, B.ID_pedido AS ID_tiempo from [dbo].[detalle_pedido] AS A
LEFT JOIN [dbo].[pedido] AS B ON A.ID_pedido = B.ID_pedido
LEFT JOIN [dbo].[cliente] AS D ON B.ID_cliente = D.ID_cliente
LEFT JOIN [dbo].[empleado] AS E ON D.ID_empleado_rep_ventas = E.ID_empleado
```

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

The screenshot shows the Microsoft SQL Server Enterprise Manager interface. The 'Query Results' window displays the execution of a query that returns data for a star schema model. The query is as follows:

```

select top (3)* from [dbo].[Cliente]
select top (3)* from [dbo].[Empleado]
select top (3)* from [dbo].[Geografia]
select top (3)* from [dbo].[Producto]
select top (3)* from [dbo].[Tiempo]
  
```

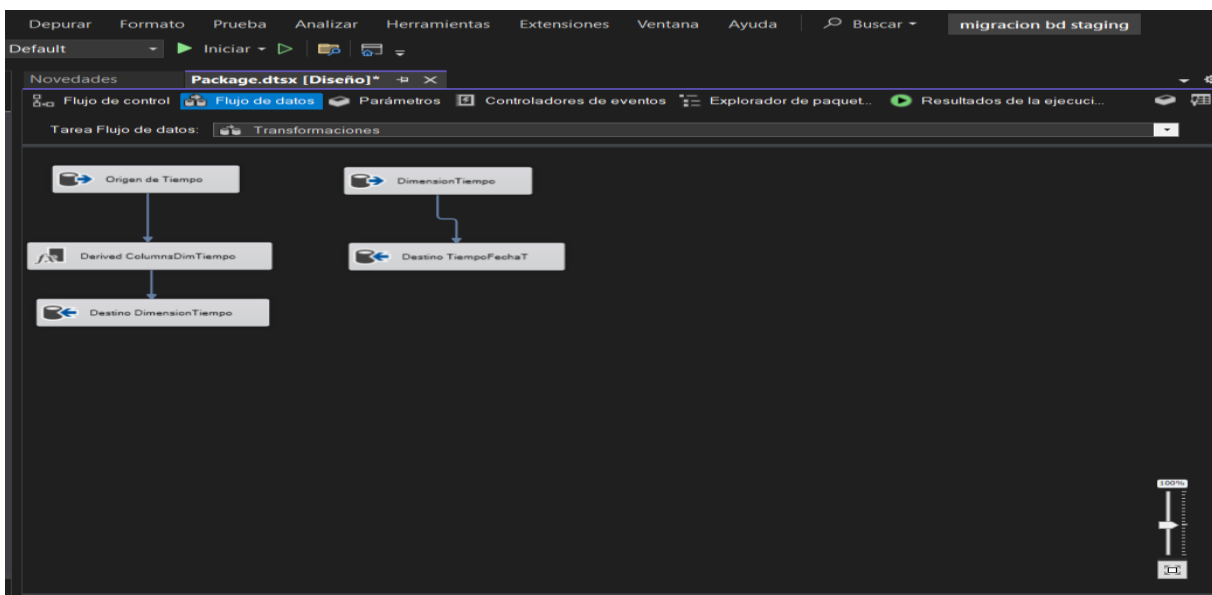
The results are displayed in a grid format, showing data for the following tables:

- Cliente:** ID_cliente, nombre_cliente, ID_empleado_res_ventas, limite_credito
- Empleado:** ID_empleado, nombre, apellido, ID_geografia
- Geografia:** ID_geografia, ciudad, pais, region
- Producto:** ID_producto, cantidad, precio_unidad, total, ID_cliente, ID_producto, ID_empleado, ID_geografia, ID_tiempo
- Tiempo:** ID_producto, codigo_producto, nombre, categoria, proveedor, cantidad_en_stock, precio_venta, precio_proveedor

The status bar at the bottom indicates 'Query executed successfully' and '18 rows'.

3. Transformación de datos según las necesidades analíticas:

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



Transformaciones

Dimension tiempo transformacion de fecha_entrega campos null

```
select [ID_tiempo],
[fecha_pedido],[fecha_esperada],
isnull(convert(varchar,
[fecha_entrega],120),0000-00-00) as
[Info_fecha_entrega],isnull([Anio],0) as
[Info_anio],isnull([Mes],0) as
[Info_mes],isnull([Dia],0) as
[Info_dia],isnull([DiasemanaNum],0) as
[Info_dia_semana_num]
from [staging].[dbo].[Destino DimensionTiempo]
```

4. 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.



```
use [staging]

select a.CodigoProducto, a.nombre, a.proveedor, SUM(isnull(a.cantidad_en_stock,0)) as Cantidad_Inventario, isnull(a.precio_proveedor,0) as Precio_Proveedor,
SUM(isnull(a.cantidad_en_stock,0)) * isnull(a.precio_proveedor,0) as Valor_Total_Inventario,
SUM(isnull(b.cantidad,0)) as Cantidad_Vendido, isnull(b.precio_unidad,0) as Valor_Unitario_Venta,
SUM(isnull(b.cantidad,0)) * isnull(b.precio_unidad,0) as Total_Venta,
isnull(b.precio_unidad,0) - isnull(a.precio_proveedor,0) as Utilidad_Venta_Unidad,
SUM(isnull(b.cantidad,0)) * (isnull(b.precio_unidad,0) - isnull(a.precio_proveedor,0)) as Utilidad_Total_Venta,
SUM(isnull(a.cantidad_en_stock,0)) * (isnull(b.precio_unidad,0) - isnull(a.precio_proveedor,0)) as Utilidad_Pendiente_Venta
from (select Distinct * from [Producto]) as a
left join (select distinct * from [dbo].[Hechos]) as b on a.ID_producto = b.ID_producto
left join (select distinct * from [dbo].[Tiempo]) as c on b.ID_tiempo = c.ID_tiempo

group by
a.CodigoProducto, a.nombre, a.proveedor, a.precio_proveedor, b.precio_unidad
```

```
use [staging]
select a.nombre + ' ' + a.apellido as Vendedor, b.pais, b.region, c.nombre_cliente as Cliente, e.nombre as Producto, SUM(d.cantidad) as Cantidad_Compra, SUM(d.total) as Total_Compra
from [dbo].[Empleado] as a
inner join (select Distinct * from [dbo].[Geografia]) as b on a.ID_geografia = b.ID_geografia
left join (select Distinct * from [dbo].[Cliente]) as c on a.ID_empleado = c.ID_empleado_rep_ventas
left join (select Distinct * from [dbo].[Hechos]) as d on c.ID_cliente = d.ID_cliente
left join (select Distinct * from [dbo].[Producto]) as e on d.ID_producto = e.ID_producto
group by
a.nombre,
a.apellido,
b.pais,
b.region,
c.nombre_cliente,
e.nombre
```

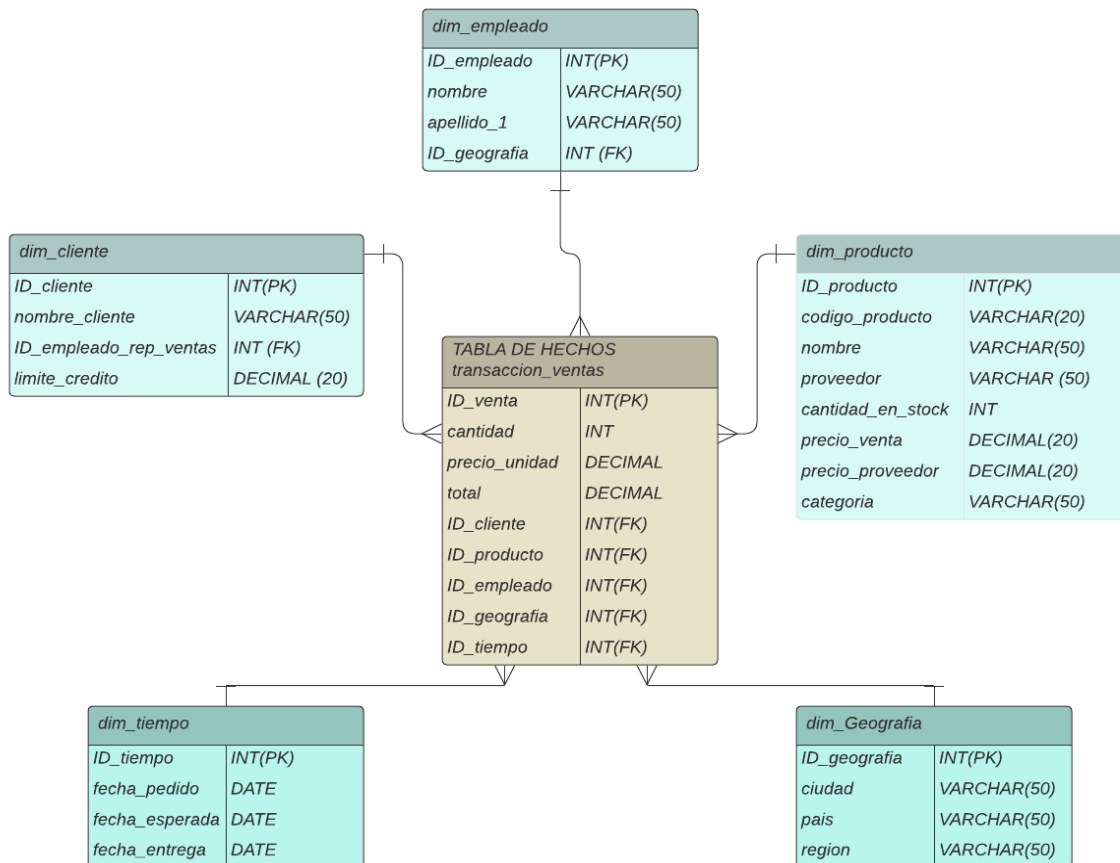
- b. Ejecutar las consultas de carga y verificar que los datos se hayan insertado correctamente en el *data mart* final.

Código/Producto	nombre	proveedor	Cantidad_Inventario	Precio_Proveedor	Valor_Total_Inventario	Cantidad_Vendida	Valor_Unidades_Ventas	Total_Ventas	Utilidad_Ventas_Unidad	Utilidad_Total_Ventas
OR-129	Camelito japonés exemplar	Viveros EL GABO	50	88.00	4400.00	1	115.00	115.00	27.00	27.00
FR-44	Melocotones Spring Crest	Melocotoneros de Cieza S.A	400	6.00	2400.00	0	0.00	0.00	-6.00	0.00
OR-299	Arachisathemum Curvigenumense	Viveros EL GABO	80	8.00	640.00	50	10.00	500.00	2.00	100.00
FR-40	Melocotones Sparkling Delicious	Frutales Talavera S.A	400	6.00	2400.00	42	9.00	378.00	3.00	126.00
OR-236	Jadwara Chileana	Viveros EL GABO	100	39.00	3900.00	45	49.00	2205.00	19.00	450.00

Vendedor	pais	region	Cliente	Producto	Cantidad_Compra	Total_Compras
Julian Ballistreri	Australia	APAC	Viveros Humanos	Thymus Citriodora (Tomillo Indio)	30	30.00
Karen Palmer	Australia	APAC	NULL	NULL	NULL	NULL
Larry Westfall	Inglaterra	EMEA	NULL	NULL	NULL	NULL
Laurent Berni	Francia	EMEA	NULL	NULL	NULL	NULL
Lionel Navarro	Francia	EMEA	francois leonard	NULL	NULL	NULL

5. Documentación y presentación:

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



Preparación

El primer paso en el desarrollo del proceso de transformación de datos y carga en el Data Mart final fue revisar el modelo Estrella definido en el proyecto de análisis de la base de datos Jardinería. Este modelo consiste en una tabla de hechos central, transaccion_ventas y sus dimensiones dim_producto, dim_empleado, dim_geografia, dim_tiempo, y dim_cliente. Este diseño permite una organización eficiente y un acceso rápido a los datos necesarios para el análisis.

Además, se verificó la disponibilidad y consistencia de la base de datos de Staging. Esta base de datos sirve como área temporal para la transformación de datos antes de su carga final en el Data Mart. Se realizaron verificaciones para asegurar que los datos en Staging estuvieran completos y correctamente relacionados, sin duplicados ni valores nulos.

Extracción de Datos

Se utilizaron consultas SQL para extraer datos relevantes de la base de datos origen hacia la base de datos de Staging, los datos extraídos incluyeron información sobre ventas, productos, empleados, clientes, y datos geográficos. Las consultas se diseñaron para tener todos los datos necesarios para alimentar el modelo Estrella y a su vez estuvieran presentes en la base de datos de Staging.

Transformación de Datos

En la etapa de transformación, se aplicaron varias técnicas para preparar los datos según las necesidades analíticas:

- ✓ **Limpieza:** Se eliminaron valores nulos y duplicados, y se corrigieron valores incorrectos.
- ✓ **Normalización:** Se aseguraron formatos consistentes para todos los campos.
- ✓ **Enriquecimiento:** Se derivaron nuevos campos a partir de los datos existentes para proporcionar un contexto más rico para el análisis.

Carga de Registros en el Data Mart Final

Se diseñaron y ejecutaron consultas SQL para cargar los datos transformados desde la base de datos de Staging en el Data Mart final, las consultas realizadas incluyeron:

- ✓ **Ventas de los vendedores en sucursales y regiones:** Carga de datos que relacionan ventas con vendedores, sucursales y regiones geográficas.
- ✓ **Clientes con más compras:** Identificación y carga de datos de clientes con el mayor número de compras.
- ✓ **Costo de productos a proveedores y ventas:** Registro del valor de compra de productos a proveedores y las ventas realizadas.
- ✓ **Margen de ganancia sobre las ventas:** Cálculo y carga de márgenes de ganancia tanto generales como específicos por venta.

Se verificó que los datos se hubieran insertado correctamente en el Data Mart final, realizando comprobaciones para asegurar la integridad y precisión de los datos.

Este informe documenta de manera clara y concisa el proceso de transformación y carga de datos, destacando las consultas SQL utilizadas y haciendo referencia al modelo Estrella definido. Esto asegura que los datos en el Data Mart final estén preparados para un análisis detallado y preciso, respaldando la toma de decisiones.

Referencias

Datamanagement. (2020, 4 junio). *Proceso ETL con SQL Server Integration Services – SSIS*, <https://datamanagement.es/2020/04/06/proceso-etl-con-sql-server-integration-services-ssis/>

Baldizzoni. E. (2013). *Propuesta de Proceso de Transformación de Datos para Proyectos de Explotación de Información*, <https://revistas.unla.edu.ar/software/article/view/96>

Alcántara. D. (2019). *Análisis, diseño e implementación de data mart de ventas para optimizar la toma de decisiones en una mediana empresa en la ciudad de Lima*, <https://repositorio.utp.edu.pe/handle/20.500.12867/2101>