

BASES DE DATOS II

POR:

SEBASTIAN BEDOYA GIRALDO

GRUPO:

PREICA2401B010093

ACTIVIDAD 3

EA3: PROCESO DE TRANSFORMACIÓN DE DATOS Y CARGA EN EL DATA MART FINAL

PROFESOR:

VICTOR HUGO MERCADO RAMOS

INSTITUCIÓN UNIVERSITARIA DIGITAL DE ANTIOQUIA

2024

Introducción

En la actualidad, el manejo de información se ha vuelto fundamental para impulsar decisiones estratégicas y alcanzar un funcionamiento óptimo. En este contexto, resulta indispensable la creación de una base de datos staging con el fin de asegurar que los datos estén completos, sean confiables y estén disponibles para su posterior análisis.

El proceso empieza al revisar el modelo estrella que se ha definido para el proyecto de análisis de la base de datos, donde se incluyen tanto la estructura como las relaciones entre las tablas dimensionales y la tabla factual. El Data mart final utiliza la base de datos de staging, que se ha creado previamente, como una importante área temporal para transformar y limpiar los datos extraídos antes de ser cargados.

Se proporcionan a continuación las consultas SQL necesarias para extraer los datos, asegurando su integridad y coherencia. Después, se implementan métodos de transformación que incluyen la limpieza y normalización de los datos. Por último, se procede a ejecutar las consultas de carga para insertar los registros transformados en el Data mart final y se verifica que los datos hayan sido correctamente ingresados.

Objetivos

1. Aplicar técnicas de transformación de datos: Realizar la limpieza, normalización y enriquecimiento de los datos en la base de datos de staging. La finalidad de estas transformaciones es mejorar la calidad de los datos y garantizar su adecuación a las necesidades del modelo principal, todo ello con el propósito de contar con información lista para ser utilizada en el Data Mart.
2. Diseñar y ejecutar consultas SQL o scripts de carga: Con el fin de llevar los datos transformados desde la base de datos de staging al Data Mart final.
3. Incorporar información adicional relevante: A medida que se van creando nuevas tablas en el modelo estrella.

Planteamiento del problema

El problema central radica en cómo transformar y cargar estos datos de manera eficiente y efectiva en un Data Mart, asegurando que estén listos para un análisis detallado y relevante. Los datos provienen de múltiples sistemas origen, cada uno con diferentes estructuras y formatos. Los datos deben ser transformados y enriquecidos para cumplir con los requisitos del modelo estrella. Para abordar estos desafíos, es esencial establecer un proceso robusto de ETL (Extract, Transform, Load) que incluya la extracción de datos de las bases de datos, asegurando que se cubren todos los aspectos necesarios para el análisis, la aplicación de técnicas de limpieza, normalización, enriquecimiento y la transferencia de los datos transformados desde la base de datos de staging al Data Mart final.

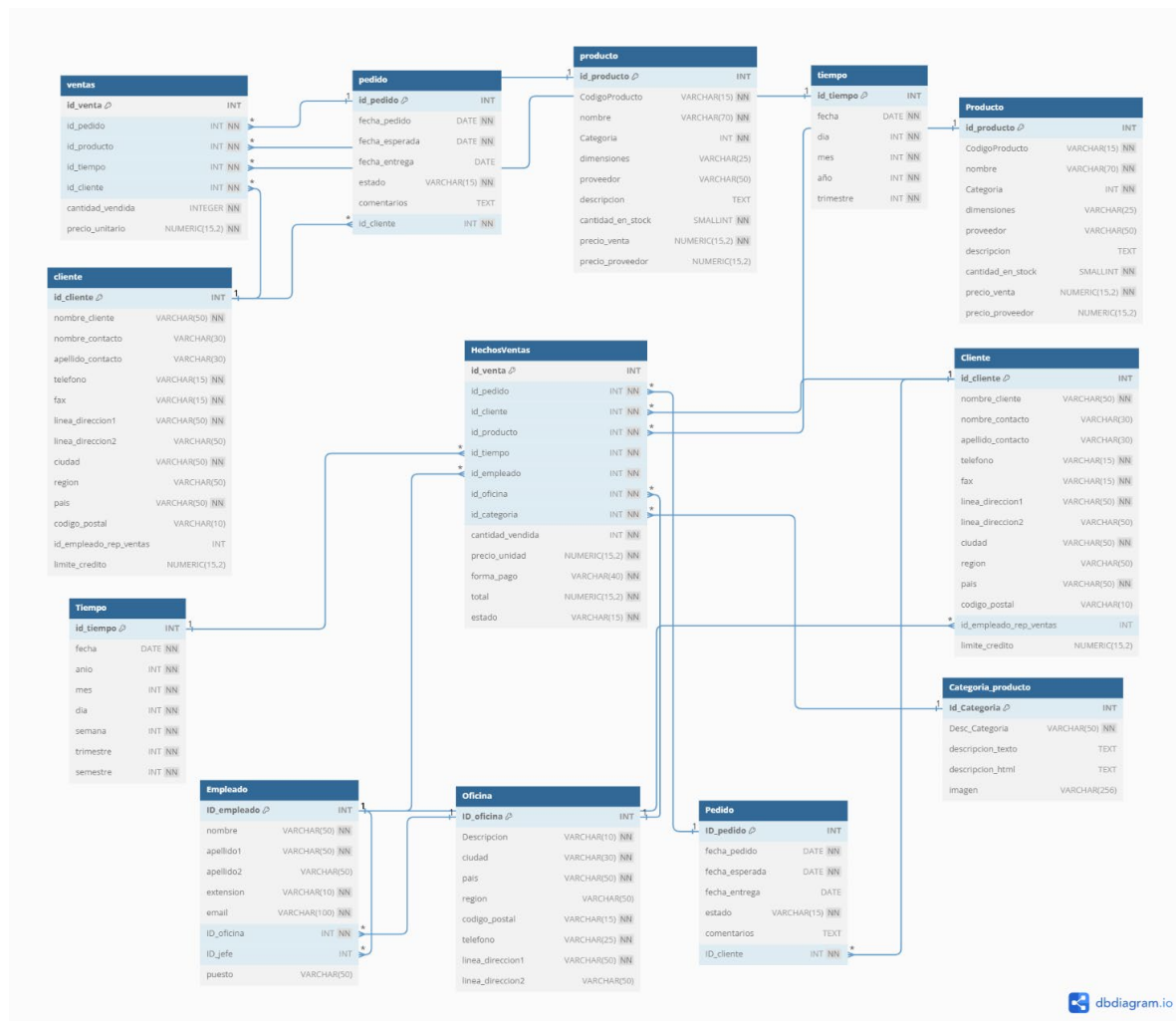
Análisis del problema

Para llevar a cabo las transformaciones, se emplean consultas SQL junto con herramientas de ETL. Mientras las consultas SQL posibilitan llevar a cabo operaciones de limpieza y transformación directamente en la base de datos de staging, las herramientas ETL son responsables tanto del manejo como la automatización eficiente del proceso de transformación de los datos. Durante esta etapa, es importante enfrentar diversos retos como asegurar la calidad de los datos, mantener consistencia en distintas fuentes y transformaciones, así como optimizar las consultas y procesos ETL para manejar grandes cantidades de información sin que esto impacte el rendimiento. Después, es necesario transferir los datos transformados desde la base de datos de staging hacia el Data Mart final. Una forma de lograr esto es a través del diseño de consultas SQL o scripts de carga que estén optimizados para insertar los datos en el Data Mart.

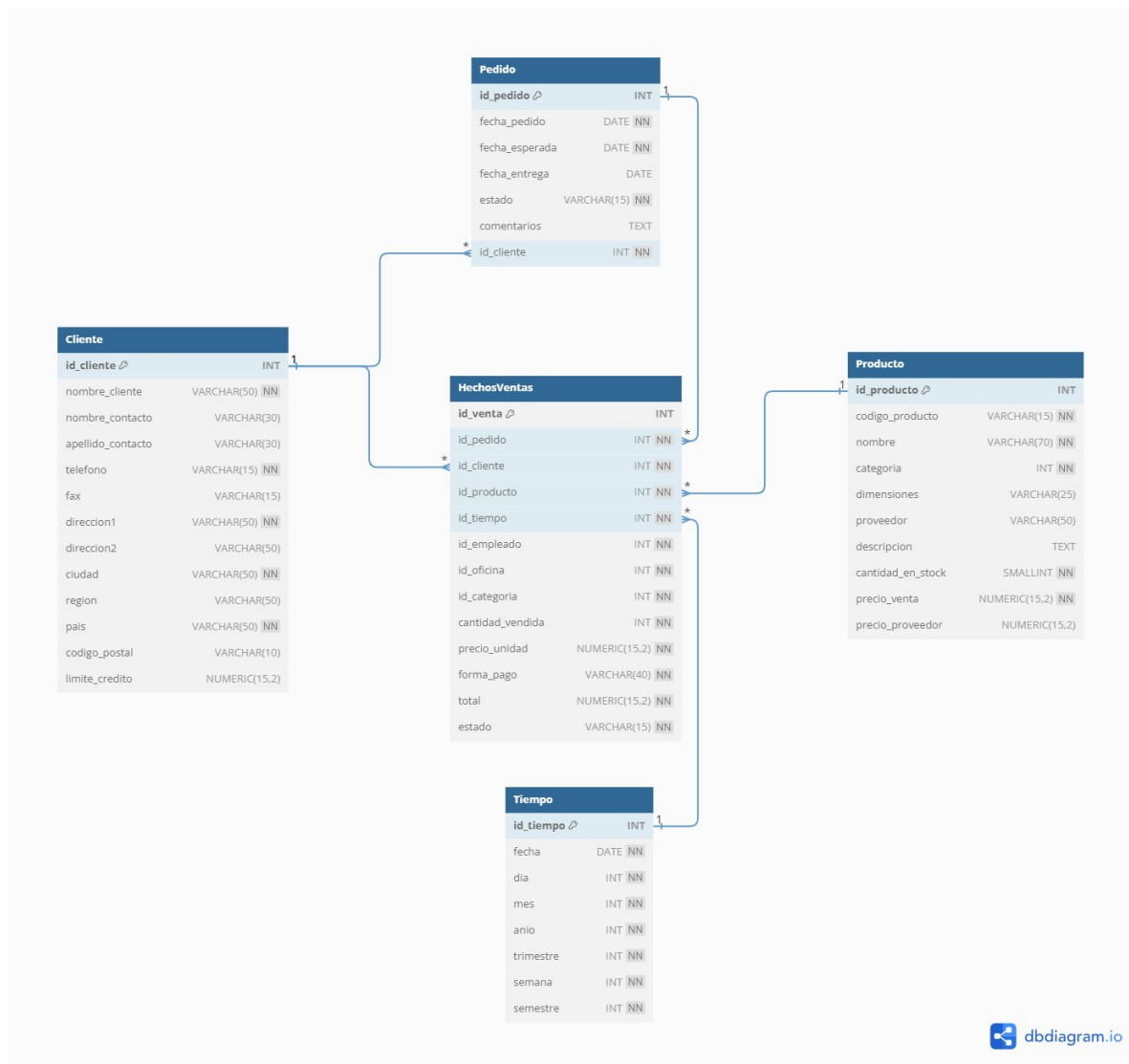
Propuesta de solución

Modelo de estrella:

MODELO ESTRELLA – PRIMERA ENTREGA

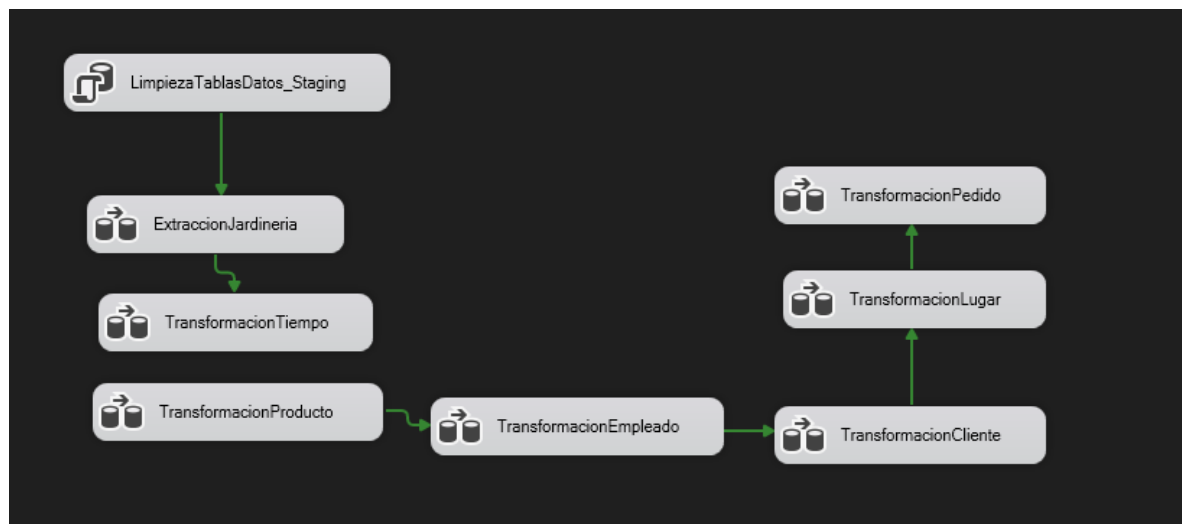


MODELO ESTRELLA – JARDINERIA CORREGIDA



Transformación de Datos según las Necesidades Analíticas

A continuación, se describe el proceso de transformación de datos para cada una de las dimensiones del modelo estrella:



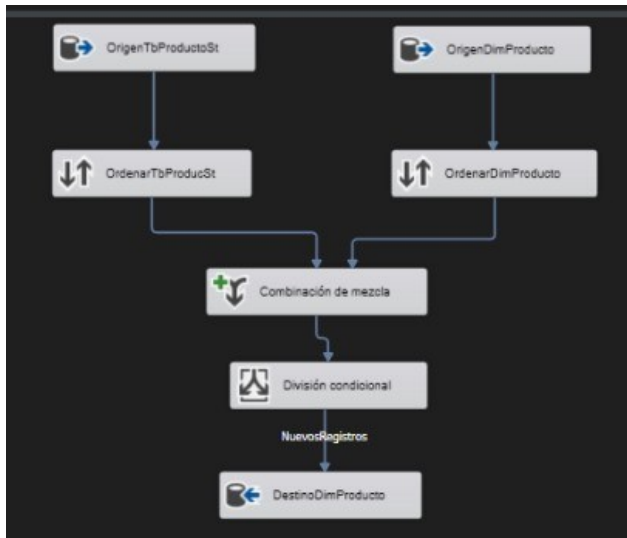
TransformacionProducto

Llamamos cada uno de los datos que harán parte de la dimensión producto de la data Mart.

```

select id_producto, id_producto_0, codigo_producto, nombre, COALESCE(NULLIF(LTRIM(RTRIM(dimensiones)), ''), 'N/A') as dimensiones,
nombre_proveedor, COALESCE(NULLIF(LTRIM(RTRIM(descripcion)), ''), 'N/A') as descripcion,
cantidad_en_stock, precio_unidad_proveedor, TbC.Desc_Categoria as nombre_categoria, TbC.descripcion_texto as descripcion_categoria
from TbProducto inner join TbCategoria TbC on Categoria = TbC.Id_Categoria_0
  
```

Esta declaración SQL se agrega al flujo de trabajo de la dimensión en ADO.NET usando OrigenTbProductoST. Antes de aplicar la combinación y mezcla con el OrigenDimProducto, se procede a ordenar estos registros por ID de producto. Luego, se realiza una división condicional para verificar si hay nuevos registros insertados utilizando la función isNull() en los IDs de productos obtenidos. DestinoDimProducto contará con la sentencia SQL para crear la tabla como dimensión (DimProducto), incluyendo todos los registros ordenados, normalizados y actualizados, además de las asignaciones predeterminadas. Se muestra de esta manera:



TransformacionEmpleado

Llamamos cada uno de los datos que harán parte de la dimensión producto de la data Mart.

```

select id_empleado, id_empleado_0, nombre,
apellido1, COALESCE(NULLIF(LTRIM(RTRIM(apellido2)), ''), 'N/A') as apellido2, extension, email,
Tb0.Descripcion as oficina, puesto from TbEmpleado TbE inner join TbOficina Tb0 on TbE.ID_oficina = Tb0.ID_oficina_0
  
```

Esta consulta en SQL forma parte de un proceso dentro del flujo de trabajo de la dimensión en ADO. Se genera la tabla DimEmpleado en DestinoDimEmpleado al combinar los registros de OrigenTbEmpleadoST, ordenándolos por ID de empleado y actualizándolos según corresponda con las asignaciones predeterminadas. Este proceso garantiza la consistencia y actualización de los datos en la base de datos junto con la información obtenida desde OrigenDimEmpleado.

Se muestra de esta manera:



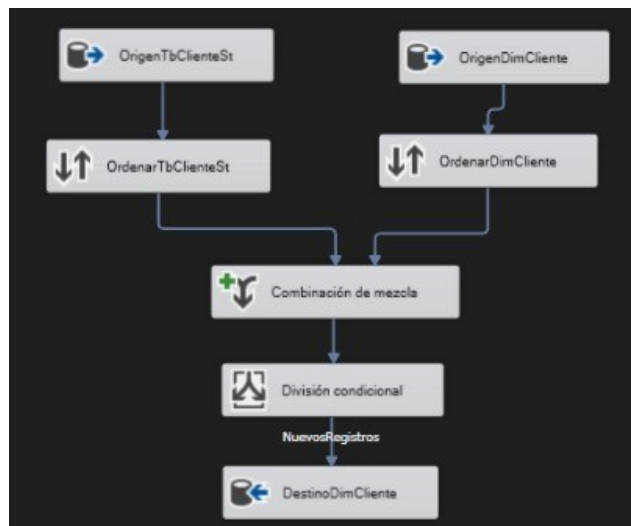
TransformacionCliente

Llamamos cada uno de los datos que harán parte de la dimensión producto de la data Mart.

```
select id_cliente, id_cliente_O,  
nombre_cliente, nombre_contacto, apellido_contacto,  
telefono, fax, linea_direccion1, COALESCE(NULLIF(LTRIM(RTRIM(linea_direccion2)), ''), 'N/A') as linea_direccion2,  
ciudad, COALESCE(NULLIF(LTRIM(RTRIM(region)), ''), 'N/A') as region from TbCliente
```

Se incorpora esta sentencia SQL en el flujo de trabajo de la dimensión mediante ADO. En NET OrigenTbClienteST, se eligen los registros de la tabla TbCliente. La sentencia recopila información detallada de los clientes y se fusiona con OrigenDimCliente. A continuación, se insertan en DestinoDimCliente para mantener actualizada la información ordenada de los registros. Esto garantiza que la base de datos tenga una completa consistencia en los detalles de cada cliente.

Se muestra de esta manera:



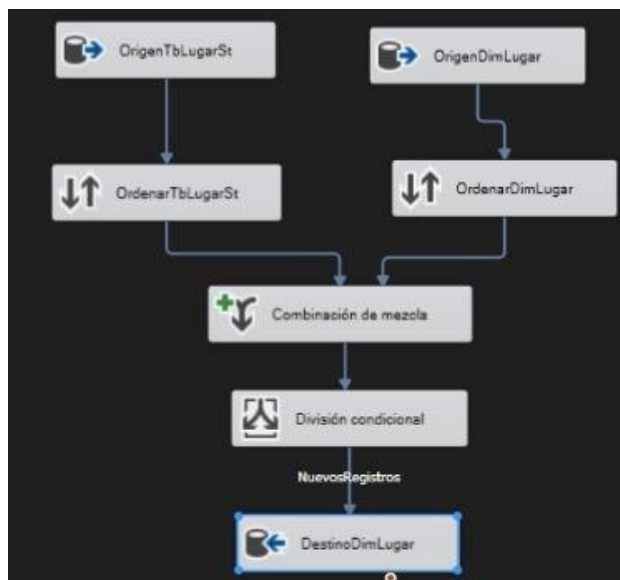
TransformaciónLugar

A continuación realizamos un llamado de cada uno de los datos que hacen parte de la dimensión Lugar en el DataMart, en este caso son: los campos de id_lugar, pais, region, ciudad, linea_direccion1.

```
select id_lugar, pais, region, ciudad, linea_direccion1 as direccion from TbLugar
```

La sentencia presentada se inserta dentro del flujo de trabajo de la dimensión **OrigenTbLugarSt**, en la cual los registros se ordenan por ID de lugar y se realiza la combinación con **OrigenDimLugar**, la consulta AS se está utilizando para renombrar la columna linea_direccion1 como dirección en el resultado de la consulta, esto quiere decir que los datos que se extraen dentro de la columna linea_direccion1 se mostrarán bajo la columna dirección.

Se muestra de esta manera:



TransformaciónTiempo

Se muestra de esta manera:



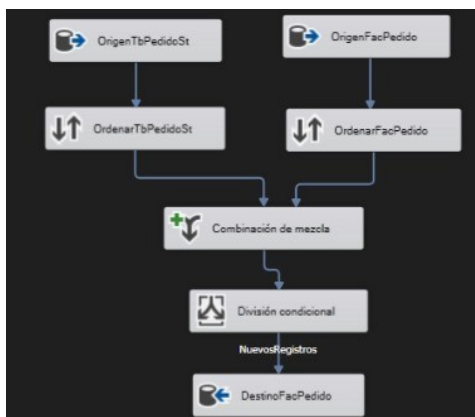
TransformacionPedido

Llamamos cada uno de los datos que harán parte de la dimensión producto escogido como Fact de la data Mart.

```
select TbP.id_pedido, id_pedido_0, DC.id_cliente, DP.id_producto, DT.id_tiempo, DE.id_empleado, DL.id_lugar,
TbD.cantidad, TbD.precio_unidad, TP.forma_pago, TP.total, estado
from TbPedido TbP inner join TbDetallePedido TbD on TbP.id_pedido_0 = TbD.ID_pedido
inner join DimProducto DP on TbD.ID_producto = DP.id_producto_0 inner join DimCliente DC on TbP.ID_cliente = DC.id_cliente_0
inner join TbCliente TbC on TbC.id_cliente_0 = TbP.ID_cliente
inner join DimTiempo DT on DT.fecha_pedido = TbP.fecha_pedido
inner join DimEmpleado DE on TbC.ID_empleado_rep_ventas = DE.id_empleado_0 inner join TbEmpleado TbE on TbE.id_empleado_0 = DE.id_empleado_0
inner join TbOficina TbO on TbO.ID_oficina_0 = TbE.ID_oficina inner join DimLugar DL on DL.direccion = TbO.linea_direccion1
inner join TbPago TP on TP.ID_cliente = DC.id_cliente_0
```

Al combinar datos de múltiples tablas relacionadas, esta consulta SQL recupera información detallada sobre los pedidos. La opción INNER JOIN se emplea para combinar la información basándose en campos de identificación compartidos, por ejemplo, el número de pedido, cliente, producto, fecha y hora del pedido, empleado a cargo y ubicación. El sistema brinda una visión detallada de cada pedido, mostrando información sobre la cantidad solicitada, el precio unitario del producto elegido por el cliente y la forma de pago utilizada.

Se muestra de esta manera:



Carga de registros en el Data Mart final:

Evidencia de los datos transformados en la dimensión Producto:

```
/****** Script para el comando SelectTopNRows de SSMS *****/
SELECT TOP (1000) [id_producto]
, [id_producto_0]
, [codigo_producto]
, [nombre]
, [dimensiones]
, [nombre_proveedor]
, [descripcion]
, [cantidad_en_stock]
, [precio_unidad_proveedor]
, [nombre_categoria]
, [descripcion_categoria]
FROM [Staging-Jardineria].[dbo].[DimProducto]
```

	codigo_producto	nombre	dimensiones	nombre_proveedor	descripcion	cantidad_en_stock	precio_unidad_proveedor	nombre_categoria	descripcion_categoria
172	OR-138	Camelia Blanco, Chrysler Rojo...	N/A	Viveros EL OASIS	N/A	100	3.00	Omamentales	Plantas vistosas para la decoración del
173	OR-139	Landora Amarillo, Rose Gaujar...	N/A	Viveros EL OASIS	N/A	100	3.00	Omamentales	Plantas vistosas para la decoración del
174	OR-140	Kordes Perfect bicolor rojo-am...	N/A	Viveros EL OASIS	N/A	100	3.00	Omamentales	Plantas vistosas para la decoración del
175	OR-141	Pitmini rojo	N/A	Viveros EL OASIS	N/A	100	3.00	Omamentales	Plantas vistosas para la decoración del
176	OR-142	Solanum Jazminoides	150-160	Viveros EL OASIS	N/A	100	1.00	Omamentales	Plantas vistosas para la decoración del
177	OR-143	Wisteria Sinensis azul, rosa, bl...	N/A	Viveros EL OASIS	N/A	100	7.00	Omamentales	Plantas vistosas para la decoración del
178	OR-144	Wisteria Sinensis INJERTADAS...	140-150	Viveros EL OASIS	N/A	100	9.00	Omamentales	Plantas vistosas para la decoración del
179	OR-145	Bougainvillea Sandieriana Tutor	80-100	Viveros EL OASIS	N/A	100	1.00	Omamentales	Plantas vistosas para la decoración del
180	OR-146	Bougainvillea Sandieriana Tutor	125-150	Viveros EL OASIS	N/A	100	3.00	Omamentales	Plantas vistosas para la decoración del
181	OR-157	Acer Pseudoplatanus	200-225	Viveros EL OASIS	N/A	100	8.00	Omamentales	Plantas vistosas para la decoración del
182	OR-158	Brachychiton Acerifolius	200-225	Viveros EL OASIS	N/A	100	4.00	Omamentales	Plantas vistosas para la decoración del
183	OR-159	Brachychiton Discolor	200-225	Viveros EL OASIS	N/A	100	4.00	Omamentales	Plantas vistosas para la decoración del
184	OR-160	Brachychiton Rupestris	170-200	Viveros EL OASIS	N/A	100	8.00	Omamentales	Plantas vistosas para la decoración del
185	OR-161	Cassia Corimbosa	200-225	Viveros EL OASIS	N/A	100	4.00	Omamentales	Plantas vistosas para la decoración del
186	OR-162	Cassia Corimbosa	200-225	Viveros EL OASIS	N/A	100	8.00	Omamentales	Plantas vistosas para la decoración del
187	OR-163	Chitaloa Summer Bells	200-225	Viveros EL OASIS	N/A	80	8.00	Omamentales	Plantas vistosas para la decoración del

Evidencia de los datos transformados en la dimensión Empleado:

```
/****** Script para el comando SelectTopNRows de SSMS *****/
SELECT TOP (1000) [id_empleado]
, [id_empleado_0]
, [nombre]
, [apellido1]
, [apellido2]
, [extension]
, [email]
, [oficina]
, [puesto]
FROM [Staging-Jardineria].[dbo].[DimEmpleado]
```

	id_empleado	id_empleado_0	nombre	apellido1	apellido2	extension	email	oficina	puesto
1	1	1	Marcos	Magaña	Perez	3897	marcos@jardineria.es	TAL-ES	Director General
2	2	2	Ruben	López	Martinez	2899	rioiez@jardineria.es	TAL-ES	Subdirector Marketing
3	3	3	Alberto	Soria	Carrasco	2837	asoria@jardineria.es	TAL-ES	Subdirector Ventas
4	4	4	Maria	Solis	Jerez	2847	msolis@jardineria.es	TAL-ES	Secretaria
5	5	5	Felipe	Rosas	Marquez	2844	frosas@jardineria.es	TAL-ES	Representante Ventas
6	6	6	Juan Carlos	Ortiz	Serrano	2845	cortiz@jardineria.es	TAL-ES	Representante Ventas
7	7	7	Carlos	Soria	Jimenez	2444	csoria@jardineria.es	MAD-ES	Director Oficina
8	8	8	Mariano	López	Murcia	2442	mlopez@jardineria.es	MAD-ES	Representante Ventas
9	9	9	Lucio	Campoamor	Martin	2442	lcampoamor@jardineria.es	MAD-ES	Representante Ventas
10	10	10	Hilario	Rodriguez	Huertas	2444	hrodriguez@jardineria.es	MAD-ES	Representante Ventas
11	11	11	Emmanuel	Magaña	Perez	2518	manu@jardineria.es	BCN-ES	Director Oficina
12	12	12	José Manuel	Martinez	De la Osa	2519	jmmart@hotmail.es	BCN-ES	Representante Ventas
13	13	13	David	Palma	Aceituno	2519	dpalma@jardineria.es	BCN-ES	Representante Ventas
14	14	14	Oscar	Palma	Aceituno	2519	opalma@jardineria.es	BCN-ES	Representante Ventas
15	15	15	Francois	Fignon	N/A	9981	ffignon@gardening.com	PAR-FR	Director Oficina
16	16	16	Lionel	Narvaez	N/A	9982	lnarvaez@gardening.com	PAR-FR	Representante Ventas
17	17	17	Laurent	Serra	N/A	9982	lserra@gardening.com	PAR-FR	Representante Ventas

Evidencia de los datos transformados en la dimensión Cliente:

***** Script para el comando SelectTopNRows de SSMS *****

```

SELECT TOP (1000) [id_cliente]
, [id_cliente_O]
, [nombre_cliente]
, [nombre_contacto]
, [apellido_contacto]
, [telefono]
, [fax]
, [linea_direccion1]
, [linea_direccion2]
, [ciudad]
, [region]
FROM [Staging-Jardineria].[dbo].[DimCliente]

```

100 %

Resultados Mensajes

	id_cliente	id_cliente_O	nombre_cliente	nombre_contacto	apellido_contacto	telefono	fax	linea_direccion1	linea_direccion2	ciudad	region
1	1	1	GoldFish Garden	Daniel G	GoldFish	5556901745	5556901746	False Street 52 2 A	NA	San Francisco	NA
2	2	2	Gardening Associates	Anne	Wright	5557410345	5557410346	Wall-e Avenue	NA	Miami	Miami
3	3	3	Gendo Valley	Link	Flaute	5552323129	5552323128	Oaks Avenue n°22	NA	New York	NA
4	4	4	Tendo Garden	Akane	Tendo	55591233210	55591233211	Null Street n°69	NA	Miami	NA
5	5	5	Lasas S.A.	Antonio	Lasas	34916540145	34914851312	C/Leganes 15	NA	Fuenlabrada	Madrid
6	6	6	Beragua	Jose	Bermejo	654867321	916549672	C/pintor segundo	Getafe	Madrid	Madrid
7	7	7	Club Golf Puerta del hierro	Peco	Lopez	624566810	919535678	C/inesio delgado	Madrid	Madrid	Madrid
8	8	8	Naturagua	Guillermo	Rengifo	689234750	916428956	C/majadahonda	Boadilla	Madrid	Madrid
9	9	9	DaraDistribuciones	David	Serrano	675598001	916421756	C/azores	Fuenlabrada	Madrid	Madrid
10	10	10	Madriña de riegos	Jose	Tacaño	655883045	916689215	C/Lagañas	Fuenlabrada	Madrid	Madrid
11	11	11	Lasas S.A.	Antonio	Lasas	34916540145	34914851312	C/Leganes 15	NA	Fuenlabrada	Madrid
12	12	12	Camunas Jardines S.L.	Pedro	Camunas	34914873241	34914871541	C/Virgenes 45	C/Princesas 2 11B	San Lorenzo del Escorial	Madrid
13	13	13	Dardena S.A.	Juan	Rodriguez	34912453217	34912484764	C/Nueva York 74	NA	Madrid	Madrid
14	14	14	Jardin de Flores	Javier	Villar	654865643	914538776	C/ Oña 34	NA	Madrid	Madrid
15	15	15	Flores Marivi	Maria	Rodriguez	666555444	912458657	C/Leganes24	NA	Fuenlabrada	Madrid
16	16	16	Flowers S.A	Beatriz	Fernandez	698754159	978453216	C/Luis Salquillo4	NA	Montomes del valles	Barcelona
17	17	17	Naturajardin	Victoria	Cruz	612343529	916548735	Plaza Maquilón 15	NA	Madrid	Madrid

Evidencia de los datos transformados en la dimensión Lugar:

***** Script para el comando SelectTopNRows de SSMS *****

```

SELECT TOP (1000) [id_lugar]
, [pais]
, [region]
, [ciudad]
, [direccion]
FROM [Staging-Jardineria].[dbo].[DimLugar]

```

%

Resultados Mensajes

	id_lugar	pais	region	ciudad	direccion
1	1	EEUU	CA	San Francisco	100 Market Street
2	2	EEUU	MA	Boston	1550 Court Place
3	3	Francia	EMEA	Paris	29 Rue Jouffroy d'abbans
4	4	Japón	Chiyoda-Ku	Tokyo	4-1 Kioicho
5	5	Australia	APAC	Sydney	5-11 Wentworth Avenue
6	6	Inglaterra	EMEA	Londres	52 Old Broad Street
7	7	España	Barcelona	Barcelona	Avenida Diagonal, 38
8	8	España	Madrid	Madrid	Bulevar Indalecio Prieto, 32
9	9	España	Castilla-LaMancha	Talavera de la Reina	Francisco Aguirre, 32

Evidencia de los datos transformados en la dimensión Tiempo:

Derived Column Name	Derived Column	Expression	Data Type	U
fecha	<add as new column>	(DT_WSTR,2)DAY(fecha_pedido) + "-" + (DT_...	cadena Unicode [DT_...	1
año	<add as new column>	YEAR(fecha_pedido)	entero de cuatro byt...	
mes	<add as new column>	MONTH(fecha_pedido)	entero de cuatro byt...	
día	<add as new column>	DAY(fecha_pedido)	entero de cuatro byt...	
semana	<add as new column>	DATEPART("dw",fecha_pedido)	entero de cuatro byt...	
semestre	<add as new column>	(DATEPART("Month",fecha_pedido) + 5) / 6	entero de cuatro byt...	
bimestre	<add as new column>	(DATEPART("Month",fecha_pedido) + 1) / 2	entero de cuatro byt...	

Evidencia de los datos transformados en la dimensión Pedido:

```

/***** Script para el comando SelectTopNRows de SSMS *****/
SELECT TOP (1000) [id_pedido]
, [id_pedido_0]
, [id_cliente]
, [id_producto]
, [id_tiempo]
, [id_empleado]
, [id_lugar]
, [cantidad]
, [precio_unidad]
, [forma_pago]
, [total]
, [estado]
FROM [Staging-Jardineria].[dbo].[FacPedido]

```

100 %

Resultados Mensajes

	id_pedido	id_pedido_0	id_cliente	id_producto	id_tiempo	id_empleado	id_lugar	cantidad	precio_unidad	forma_pago	total	estado
1	1	1	5	87	1	8	8	10	70.00	Transferencia	23794.00	Entregado
2	1	1	5	151	1	8	8	40	4.00	Transferencia	23794.00	Entregado
3	1	1	5	165	1	8	8	25	4.00	Transferencia	23794.00	Entregado
4	1	1	5	265	1	8	8	15	19.00	Transferencia	23794.00	Entregado
5	1	1	5	276	1	8	8	23	14.00	Transferencia	23794.00	Entregado
6	2	2	5	271	8	8	8	5	462.00	Transferencia	23794.00	Entregado
7	2	2	5	251	8	8	8	67	64.00	Transferencia	23794.00	Entregado
8	2	2	5	183	8	8	8	12	6.00	Transferencia	23794.00	Entregado
9	2	2	5	58	8	8	8	7	8.00	Transferencia	23794.00	Entregado
10	2	2	5	164	8	8	8	50	4.00	Transferencia	23794.00	Entregado
11	2	2	5	57	8	8	8	3	29.00	Transferencia	23794.00	Entregado
12	2	2	5	165	8	8	8	20	5.00	Transferencia	23794.00	Entregado
13	3	3	5	241	14	8	8	15	65.00	Transferencia	23794.00	Rechazado
14	3	3	5	237	14	8	8	30	266.00	Transferencia	23794.00	Rechazado
15	3	3	5	242	14	8	8	24	25.00	Transferencia	23794.00	Rechazado
16	3	3	5	146	14	8	8	32	5.00	Transferencia	23794.00	Rechazado
17	3	3	5	66	14	8	8	120	9.00	Transferencia	23794.00	Rechazado

Conclusiones

- **Construcción y Integración de la Base de Datos Staging:** Es fundamental establecer y luego fusionar una base de datos staging con un Data Mart para poder administrar los datos de manera eficiente. En este proceso se realiza la revisión del modelo estrella, se extraen datos de diferentes fuentes, se aplican transformaciones para mejorar la calidad y consistencia de los datos, y por último se cargan estos datos transformados en el Data Mart. Los retos principales a enfrentar son garantizar la calidad y uniformidad de los datos, así como mejorar el procedimiento ETL para trabajar con grandes cantidades de información sin afectar el rendimiento.
- **Mejora en la Gestión de Datos:** Lograr con éxito la implementación de una base de datos staging y su integración en un Data Mart brinda una sólida plataforma para realizar análisis detallados y mejorar la gestión de datos. Este logra cumplir los objetivos planteados y establece un alto estándar para proyectos de datos futuros, garantizando que la información esté disponible y sea confiable al momento de tomar decisiones estratégicas.