

EVIDENCIA DE APRENDIZAJE 3. PROCESO DE TRANSFORMACIÓN DE DATOS
Y CARGA EN EL DATA MART FINAL

Por
WILFRAN CAMILO VALENCIA GOEZ

Materia
BASE DE DATOS II

INSTITUCIÓN UNIVERSITARIA DIGITAL DE ANTIOQUIA

MEDELLÍN

2025

Introducción

En estos momentos la gestión de datos ha cambiado en esencial para impulsar la toma de decisiones estratégicas. En esto, la construcción de una base de datos staging surge como una tarea fundamental para garantizar la integridad y disponibilidad de los datos para análisis, el proceso empieza revisando el modelo estrella definido para el proyecto de análisis de la base de datos, el cual también está la estructura y relaciones entre las tablas de dimensiones y la tabla de hechos. La base de datos creada, muestra como almacén temporal donde los datos extraídos son transformados y limpiados antes de ser cargados en el Data Mart final.

También las consultas SQL necesarias para la extracción de datos, dejando en claro que los datos son consistentes. Además, se aplican técnicas de transformación, como la limpieza y normalización de datos, finalmente, se ejecutan las consultas de carga que insertan los registros transformados en el Data Mart final, verificando que los datos se hayan insertado correctamente.

Objetivos

- Primero, vamos a arreglar y adaptar los datos en la zona de preparación. Los vamos a limpiar (quitar errores), normalizar (unificar formatos) y mejorar (añadir información) para que sean de alta calidad y encajen perfectamente en la estructura de análisis (el modelo estrella).
- Luego, usaremos código (SQL o scripts) para transferir estos datos ya listos y transformados desde la zona de preparación a la base de datos de destino (Data Mart).
- Finalmente, nos aseguraremos de incluir cualquier dato nuevo o tabla extra que sea importante para hacer los análisis completos.

Planteamiento Del Problema

El desafío es preparar los datos que provienen de sistemas origen heterogéneos para cargarlos en el Data Mart, listos para un análisis detallado.

Para resolverlo, es imprescindible implementar un proceso ETL completo. Este proceso debe extraer todos los datos necesarios, transformarlos mediante limpieza, normalización y enriquecimiento para alinearlos con el modelo estrella, y luego cargarlos eficientemente desde el área de staging al destino final.

Análisis Del Problema

Para ejecutar las transformaciones, se emplean consultas SQL que operan directamente sobre la base de staging, junto con herramientas ETL que ofrecen una gestión y automatización eficiente.

El principal objetivo es garantizar la calidad y consistencia de los datos a lo largo de las transformaciones. Sin embargo, el gran desafío es optimizar tanto las consultas como los procesos ETL para mantener un buen rendimiento al trabajar con grandes volúmenes de información.

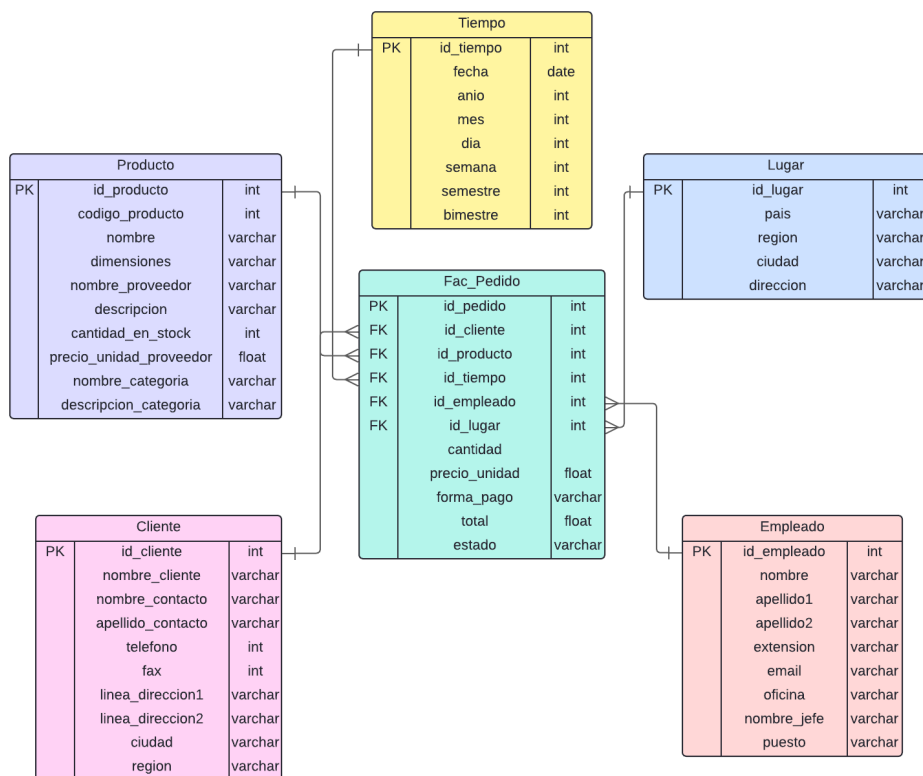
Una vez transformados, el paso final es mover los datos del área de staging al Data Mart, utilizando consultas SQL o scripts de carga diseñados específicamente para una inserción rápida y eficiente.

Propuesta de la solución con

Modelo estrella - primera entrega

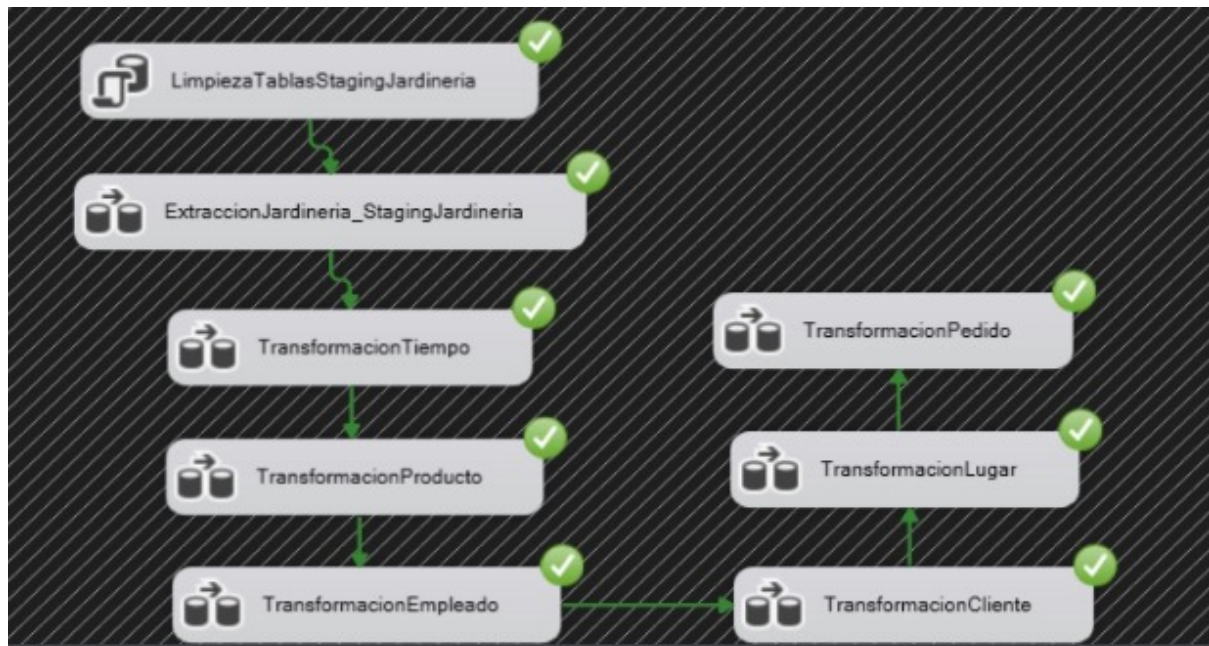


Modelo estrella - Jardinería corregido



Transformación de datos según las necesidades analíticas:

A partir de este punto, el flujo de trabajo de extracción da paso a flujos de trabajo paralelos, cada uno dedicado a la transformación de una de las seis dimensiones establecidas.



TransformacionProducto

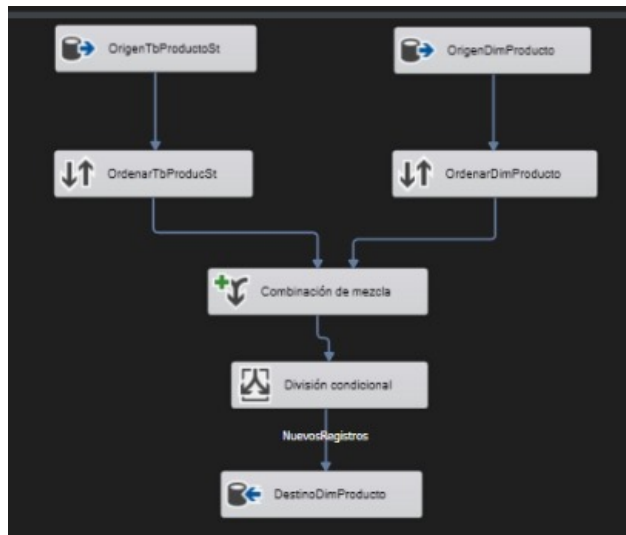
Se llama todos los datos que harán parte de la dimensión producto del data Mart.

```

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 sentencia Sql se inserta dentro del flujo de trabajo de la dimensión en ADO net **OrigenTbProductoST**. La transformación de estos registros primeramente son ordenados por ID de producto antes de aplicarle **combinación de mezcla** con **OrigenDimProducto**, posteriormente se le aplica una **división condicional** que verifica la aparición de nuevos registros insertados por medio de la función `isNull()` aplicado a los ID de los productos obtenidos. **DestinoDimProducto** tendrá la sentencia SQL para la creación de la tabla que funciona como dimensión (**DimProducto**) con los todos los registros ordenados, normalizados, actualizados y con las asignaciones predeterminadas.

Se muestra de esta manera:



TransformacionEmpleado

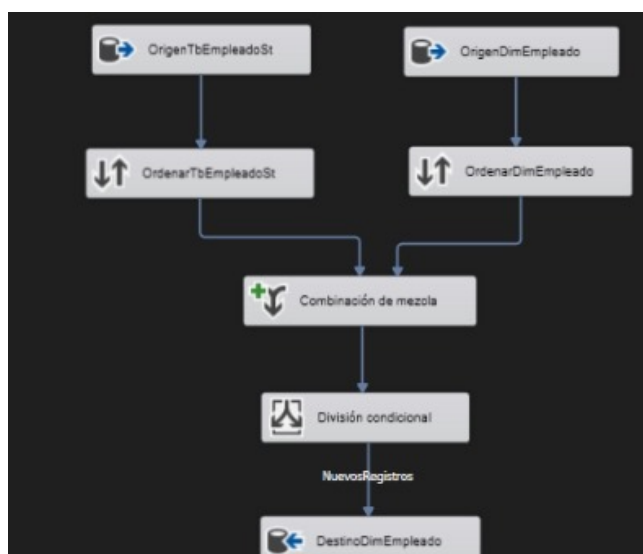
Se llama cada uno de los datos que harán parte de la dimensión producto del data Mart.

```

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 forma parte del proceso ADO.NET que gestiona la dimensión DimEmpleado. Su objetivo es fusionar la nueva información proveniente del staging (OrigenTbEmpleadoST, previamente ordenada por ID de empleado) con los datos históricos (OrigenDimEmpleado). El resultado de esta combinación se usa para actualizar la tabla DimEmpleado (DestinoDimEmpleado), asegurando que los registros estén ordenados, actualizados y sean consistentes mediante la aplicación de valores predeterminados.

Se muestra de esta manera:



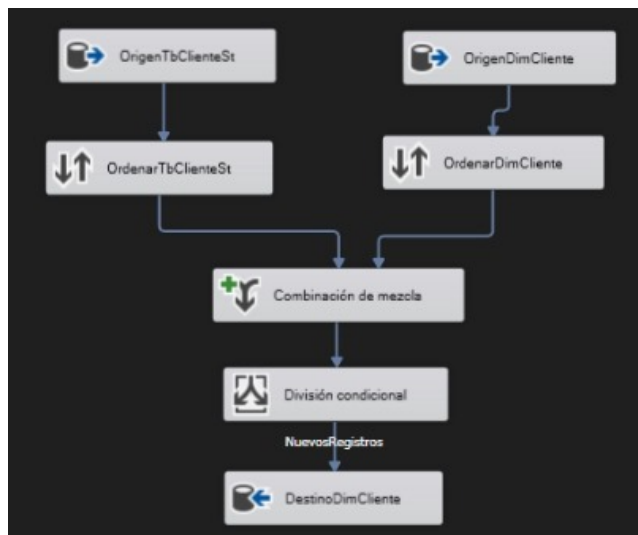
TransformacionCliente

Se llama los datos que harán parte de la dimensión producto del 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
```

Esta sentencia SQL es el motor de actualización de la dimensión Cliente en el flujo de trabajo ADO.NET. La lógica consiste en extraer los detalles del cliente de la tabla TbCliente, combinarlos con los datos existentes (OrigenDimCliente) y luego insertar el conjunto de registros actualizado y ordenado en la tabla final de destino (DestinoDimCliente).

Se muestra de esta manera:



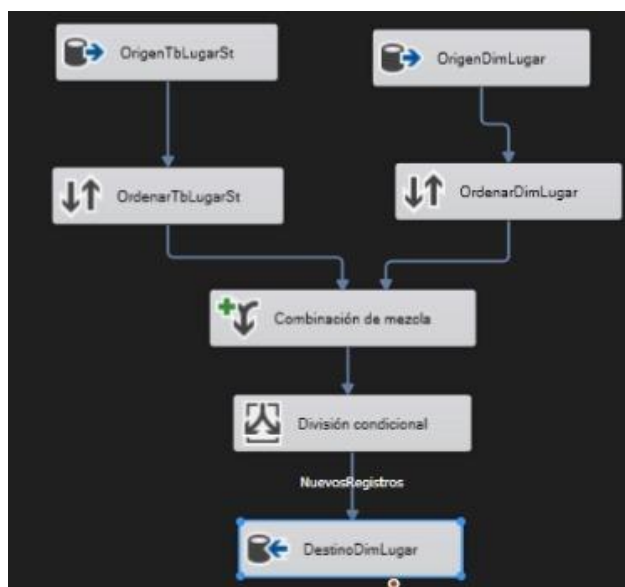
TransformaciónLugar

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
```

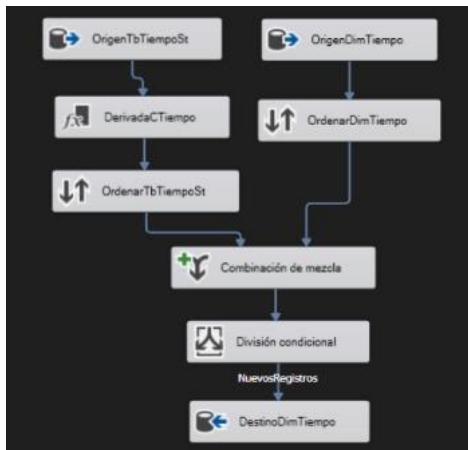
La sentencia en el flujo de trabajo de la dimensión Lugar (OrigenTbLugarSt) tiene dos funciones: ordenar los datos por ID y combinarlos con OrigenDimLugar. Adicionalmente, usa la cláusula AS para cambiar el nombre de la columna linea_direccion1 a dirección en el resultado de la consulta.

Se muestra de esta manera:



TransformaciónTiempo

Se muestra de esta manera:



TransformacionPedido

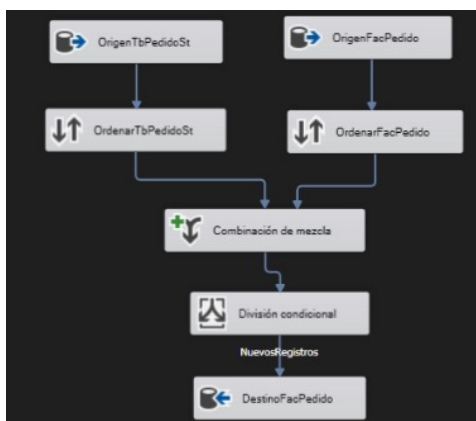
Se llama los datos que harán parte de la dimensión producto escogido como Fact del data Mart.

```

select TbP.id_pedido, id_pedido_O, 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_O = TbD.ID_pedido
inner join DimProducto DP on TbD.ID_producto = DP.id_producto_O inner join DimCliente DC on TbP.ID_cliente = DC.id_cliente_O
inner join TbCliente TbC on TbC.id_cliente_O = 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_O inner join TbEmpleado TbE on TbE.id_empleado_O = DE.id_empleado_O
inner join TbOficina TbO on TbO.ID_oficina_O = TbE.ID_oficina inner join DimLugar DL on DL.direccion = TbO.linea_direccion1
inner join TbPago TP on TP.ID_cliente = DC.id_cliente_O
  
```

La función de esta sentencia SQL es extraer todos los detalles de los pedidos al combinar múltiples tablas relacionadas (cliente, producto, empleado, etc.). La unión se realiza de forma precisa mediante INNER JOIN en identificadores comunes, generando una vista que incluye todos los datos transaccionales, como cantidad, precio, pago y estado del pedido.

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]

```

0.00 %

Resultados Mensajes

	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	Ornamentales	Plantas vistosas para la decoración del j
173	OR-139	Landora Amarillo, Rose Gaujar...	N/A	Viveros EL OASIS	N/A	100	3.00	Ornamentales	Plantas vistosas para la decoración del j
174	OR-140	Kordes Perfect bicolor rojo-am...	N/A	Viveros EL OASIS	N/A	100	3.00	Ornamentales	Plantas vistosas para la decoración del j
175	OR-141	Pitmini rojo	N/A	Viveros EL OASIS	N/A	100	3.00	Ornamentales	Plantas vistosas para la decoración del j
176	OR-142	Solanum Jazminoides	150-160	Viveros EL OASIS	N/A	100	1.00	Ornamentales	Plantas vistosas para la decoración del j
177	OR-143	Wisteria Sinensis azul, rosa, bl...	N/A	Viveros EL OASIS	N/A	100	7.00	Ornamentales	Plantas vistosas para la decoración del j
178	OR-144	Wisteria Sinensis INJERTADAS...	140-150	Viveros EL OASIS	N/A	100	9.00	Ornamentales	Plantas vistosas para la decoración del j
179	OR-145	Bougamvillea Sanderiana Tutor	80-100	Viveros EL OASIS	N/A	100	1.00	Ornamentales	Plantas vistosas para la decoración del j
180	OR-146	Bougamvillea Sanderiana Tutor	125-150	Viveros EL OASIS	N/A	100	3.00	Ornamentales	Plantas vistosas para la decoración del j
181	OR-157	Acer Pseudoplatanus	200-225	Viveros EL OASIS	N/A	100	8.00	Ornamentales	Plantas vistosas para la decoración del j
182	OR-158	Brachychiton Acerifolius	200-225	Viveros EL OASIS	N/A	100	4.00	Ornamentales	Plantas vistosas para la decoración del j
183	OR-159	Brachychiton Discolor	200-225	Viveros EL OASIS	N/A	100	4.00	Ornamentales	Plantas vistosas para la decoración del j
184	OR-160	Brachychiton Rupestris	170-200	Viveros EL OASIS	N/A	100	8.00	Ornamentales	Plantas vistosas para la decoración del j
185	OR-161	Cassia Corimbosa	200-225	Viveros EL OASIS	N/A	100	4.00	Ornamentales	Plantas vistosas para la decoración del j
186	OR-162	Cassia Corimbosa	200-225	Viveros EL OASIS	N/A	100	8.00	Ornamentales	Plantas vistosas para la decoración del j
187	OR-163	Chitalos Summer Bells	200-225	Viveros EL OASIS	N/A	80	8.00	Ornamentales	Plantas vistosas para la decoración del j

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]

```

%

Resultados Mensajes

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

```

	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	N/A	San Francisco	N/A
2	2	2	Gardening Associates	Anne	Wright	5557410345	5557410346	Wall-e Avenue	N/A	Miami	Miami
3	3	3	Gerudo Valley	Link	Flaute	5552323129	5552323128	Oake Avenue n22	N/A	New York	N/A
4	4	4	Tendo Garden	Akane	Tendo	55591233210	55591233211	Null Street n169	N/A	Miami	N/A
5	5	5	Lassas S.A.	Antonio	Lassas	34916540145	34914851312	CLeganes 15	N/A	Fuenlabrada	Madrid
6	6	6	Bersguia	Jose	Berniejo	654987321	916549872	C/pintor segundo	Getafe	Madrid	Madrid
7	7	7	Club Golf Puerta del hieno	Paco	Lopez	62456810	919535678	C/inesio delgado	Madrid	Madrid	Madrid
8	8	8	Naturagus	Guillermo	Rengifo	689234750	916428956	C/majadahonda	Bosdilla	Madrid	Madrid
9	9	9	DadaDistribuciones	David	Serrano	675588001	916421756	C/azores	Fuenlabrada	Madrid	Madrid
10	10	10	Madridiña de riegos	Jose	Tacaño	655682045	916689215	CLegañaes	Fuenlabrada	Madrid	Madrid
11	11	11	Lassas S.A.	Antonio	Lassas	34916540145	34914851312	CLeganes 15	N/A	Fuenlabrada	Madrid
12	12	12	Camunas Jardines S.L.	Pedro	Camunas	34914873241	34914871541	CLeganes 45	C/Princesas 2 11B	San Lorenzo del Escorial	Madrid
13	13	13	Dardena S.A.	Juan	Rodriguez	34912453217	34912484764	CL Nueva York 74	N/A	Madrid	Madrid
14	14	14	Jardin de Flores	Javier	Villar	654865643	914538776	CL Oña 34	N/A	Madrid	Madrid
15	15	15	Flores Marivi	Maria	Rodriguez	666555444	912459657	CLeganes24	N/A	Fuenlabrada	Madrid
16	16	16	Flowers S.A	Beatriz	Fernandez	698754159	978453216	CLuis Salquillo4	N/A	Montornes del valles	Barcelona
17	17	17	Naturajardin	Victoria	Cruz	612343529	916548735	Plaza Magallón 15	N/A	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]

```

	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
fecha	<add as new column>	(DT_WSTR,2)DAY(fecha_pedido) + "-" + (DT_...	cadena Unicode [DT_...
anio	<add as new column>	YEAR(fecha_pedido)	entero de cuatro byt...
mes	<add as new column>	MONTH(fecha_pedido)	entero de cuatro byt...
dia	<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_O]
, [id_cliente]
, [id_producto]
, [id_tiempo]
, [id_empleado]
, [id_lugar]
, [cantidad]
, [precio_unidad]
, [forma_pago]
, [total]
, [estado]
FROM [Staging-Jardineria].[dbo].[FacPedido]

```

	id_pedido	id_pedido_O	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

Conclusión

- La integración de la base de datos staging y el Data Mart es crítica. El proceso esencialmente revisa el modelo de análisis, extrae datos heterogéneos, aplica transformaciones de calidad y carga el resultado en el Data Mart. El gran desafío radica en mantener la consistencia y optimizar el proceso ETL para manejar grandes volúmenes. Completar esta integración con éxito es clave, ya que establece una base sólida para análisis detallados y una gestión de datos superior.