

Proceso de transformación de datos y carga en el data mart final

Kevin Andrés Argumedo Cadavid

Astrid Carolina Hernández Torres

Institución Universitaria Digital de Antioquia

Facultad de ingeniería

Ingeniería de software y datos

Base de datos II

Antonio Jesús Valderrama Jaramillo

Urabá Antioquia

28/09/2025

Link Archivos

<https://drive.google.com/drive/folders/1VYNkS7efuiNe-BVwD1TSj1PYCw0wCQET?usp=sharing>

Extracción de datos

A Través de las sentencias insert into extraemos desde **jardineria1** hacia las tablas staging solo la información relevante según las condiciones de negocio de nuestro data mart.

Limpiamos tabla productos staging

```
TRUNCATE TABLE stg_producto;
```

Cargar información staging producto

```
insert into stg_producto (
    ID_producto ,
    CodigoProducto,
    Nombre,
    Categoria,
    Dimensiones,
    Proveedor
) SELECT
    p.ID_producto,
    p.codigoProducto,
    LTRIM(RTRIM(p.nombre)) AS nombre,
    p.Categoria,
    p.dimensiones,
    p.proveedor
FROM jardineria1.dbo.producto p
WHERE p.CodigoProducto is not null
AND p.nombre is not null
AND p.precio_proveedor >0
AND p.precio_venta >0
AND p.cantidad_en_stock >=1;
```

Limpiamos tabla Staging Fecha

```
TRUNCATE TABLE stg_fecha;
```

Cargar informacion staging fecha

```
INSERT INTO stg_fecha (
    Fecha,
    Anio,
    Mes,
    Dia
)
SELECT DISTINCT
    CAST(fecha_pedido AS DATE ) AS Fecha,
    YEAR(fecha_pedido) AS Anio,
    MONTH(fecha_pedido) AS Mes,
    DAY(fecha_pedido) AS Dia
FROM jardineria1.dbo.pedido p
WHERE p.fecha_pedido is not null
and p.fecha_entrega is not null ;
```

Limpiamos tabla Staging Ventas

```
TRUNCATE TABLE stg_ventas;
```

Cargar informacion staging ventas

```
INSERT INTO stg_ventas (
    id_stg_producto,
    id_stg_fecha,
    Cantidad_vendida,
    Precio_unidad,
    Total_venta
)
SELECT
    sp.id_stg_producto,
    sf.id_stg_fecha,
    dp.cantidad,
    dp.precio_unidad,
    dp.cantidad * dp.precio_unidad AS Total_venta
FROM jardinerial.dbo.detalle_pedido dp
JOIN jardinerial.dbo.pedido p on dp.ID_pedido = p.ID_pedido
JOIN stg_producto sp on dp.ID_producto = sp.ID_producto
JOIN stg_fecha sf on CAST (p.fecha_pedido AS DATE) = sf.Fecha
WHERE dp.precio_unidad is not null and dp.cantidad >0;
```

Limpieza de datos

Realizamos limpieza identificando y reportando registros inválidos y eliminando duplicados en todas las tablas de staging, aplicando reglas para campos obligatorios, y formatos. Esto asegura que los datos en staging cumplen normas de integridad y están limpios para ser cargados al Data Mart final.

Filtrado de valores Inválidos

```
SELECT 'Códigos inválidos' AS problema, COUNT(*) AS cantidad
FROM stg_producto
WHERE CodigoProducto IS NULL
    OR LEN(CodigoProducto)<3
    OR LEN (CodigoProducto)>15
    OR PATINDEX('%[^A-Za-z0-9-]% ', CodigoProducto)>0
UNION ALL
SELECT 'Duplicados productos', COUNT(*)
FROM (
    SELECT ID_producto
    FROM stg_producto
    GROUP BY ID_producto
    HAVING COUNT(*) > 1) d
UNION ALL
SELECT 'Duplicados fechas', COUNT(*)
FROM (
    SELECT Fecha
    FROM stg_fecha
    GROUP BY Fecha
    HAVING COUNT(*) > 1) d;
```

Eliminación de Valores inválidos

```
DELETE p1
FROM stg_producto p1
INNER JOIN stg_producto p2
ON p1.ID_producto = p2.ID_producto
WHERE p1.id_stg_producto < p2.id_stg_producto;
DELETE f1
FROM stg_fecha f1
INNER JOIN stg_fecha f2
ON f1.Fecha = f2.Fecha
WHERE f1.id_stg_fecha > f2.id_stg_fecha;
```

Transformación de Datos

Se toman los valores que se encuentran vacíos de la base datos original, tabla productos columna dimensiones, dichos valores vacíos pasan a ser “No especificados” para llevar un mejor control.

Normalización de valores vacíos en columna Dimensiones

```
UPDATE stg_producto
SET Dimensiones = 'No especificado'
WHERE Dimensiones IS NULL OR LTRIM(RTRIM(Dimensiones)) = '';
```

Detección de anomalías

Detectar productos con precios incoherentes y ayudar a corregir la fuente de los datos o hacer reportes de calidad documental.

Identificación de anomalías en precios de proveedor

```
SELECT ID_producto, precio_proveedor, precio_venta
FROM jardineria1.dbo.producto
WHERE precio_proveedor <= 0 OR precio_venta <= 0 OR precio_venta <
precio_proveedor
```

Carga de Datos a DataMart

```
INSERT INTO DataMartJardineria.dbo.dim_producto (
    ID_producto,
    CodigoProducto,
    Nombre,
    Categoria,
    Dimensiones,
    Proveedor)
SELECT ID_producto, CodigoProducto, Nombre, Categoria, Dimensiones,
Proveedor
FROM StagingJardineria.dbo.stg_producto;

INSERT INTO DataMartJardineria.dbo.dim_fecha (
    fecha_key,
    fecha,
    anio,
    mes,
    dia)
SELECT id_stg_fecha, Fecha, Anio, Mes, Dia
FROM stg_fecha;

INSERT INTO DataMartJardineria.dbo.Fact_venta (
    Product_key,
    Fecha_key,
    cantidad_vendida,
    precio_unidad,
    total_venta
)
SELECT
    dp.Product_key,          -- surrogate key de producto en la dimensión
    df.Fecha_key,            -- surrogate key de fecha en la dimensión
    sv.Cantidad_vendida,
    sv.Precio_unidad,
    sv.Total_venta
FROM StagingJardineria.dbo.stg_ventas sv
JOIN StagingJardineria.dbo.stg_producto sp
    ON sv.id_stg_producto = sp.id_stg_producto
JOIN DataMartJardineria.dbo.Dim_producto dp
    ON sp.ID_producto = dp.ID_producto
JOIN StagingJardineria.dbo.stg_fecha sf
    ON sv.id_stg_fecha = sf.id_stg_fecha
JOIN DataMartJardineria.dbo.Dim_fecha df
    ON sf.Fecha = df.Fecha;
```

Dim_producto

Productos en DataMart ya normalizados “No especificados” para los campos donde las dimensiones del producto este vacío.

15	15	25	FR-108	Peral	4	12/14	Frutales Talavera S.A
16	16	26	FR-11	Limonero 30/40	4	No especificado	NaranjasValencianas.com
17	17	27	FR-12	Kunquat	4	No especificado	NaranjasValencianas.com
18	18	28	FR-13	Kunquat EXTRA con FRUTA	4	150-170	NaranjasValencianas.com
19	19	29	FR-14	Calamondin Mini	4	No especificado	Frutales Talavera S.A
20	20	30	FR-15	Calamondin Copa	4	No especificado	Frutales Talavera S.A
21	21	31	FR-16	Calamondin Copa EXTRA Con FRUTA	4	100-120	Frutales Talavera S.A
22	22	32	FR-17	Rosal bajo 1Aº -En maceta-inicio brotación	4	No especificado	Frutales Talavera S.A
23	23	33	FR-18	ROSAL TREPADOR	4	No especificado	Frutales Talavera S.A
24	24	34	FR-19	Camelia Blanco, Chrysler Rojo, Soraya Naranja,	4	No especificado	NaranjasValencianas.com
25	25	35	FR-2	Naranja -Plantón joven 1 año injerto	4	No especificado	NaranjasValencianas.com
26	26	36	FR-20	Landra Amarillo, Rosa Gaulard hicolor blanco-r	4	No especificado	Frutales Talavera S.A
248	248	259	OR-235	Dracaena Drago	5	150 - 175	
249	249	260	OR-236	Jubaea Chilensis	5	No especificado	
250	250	261	OR-237	Livistonia Australis	5	100 - 125	
251	251	262	OR-238	Livistonia Decipiens	5	90 - 110	
252	252	263	OR-239	Livistonia Decipiens	5	180 - 200	
253	253	264	OR-240	Phoenix Canariensis	5	110 - 130	
254	254	265	OR-241	Phoenix Canariensis	5	180 - 200	
255	255	266	OR-242	Rhaphis Excelsa	5	80 - 100	
256	256	267	OR-243	Rhaphis Humilis	5	150- 170	
257	257	268	OR-244	Sabal Minor	5	60 - 75	
258	258	269	OR-245	Sabal Minor	5	120 - 140	
259	259	270	OR-246	Trachycarpus Fortunei	5	90 - 105	
260	260	271	OR-247	Trachycarpus Fortunei	5	250-300	
261	261	272	OR-248	Washingtonia Robusta	5	60 - 70	
262	262	273	OR-249	Washingtonia Robusta	5	130 - 150	
263	263	274	OR-250	Yucca Jewel	5	80 - 105	
264	264	275	OR-251	Zamia Furfuracea	5	90 - 110	
265	265	276	OR-99	Mimosa DEALBATA Gaulois Astier	5	200-225	

Dim_Fecha

	123 Fecha key	fecha	123 anio	123 mes	123 dia
3	3	2007-01-07	2,007	1	7
4	4	2007-03-19	2,007	3	19
5	5	2007-06-20	2,007	6	20
6	6	2007-10-05	2,007	10	5
7	7	2007-10-23	2,007	10	23
8	8	2008-01-04	2,008	1	4
9	9	2008-03-05	2,008	3	5
10	10	2008-03-17	2,008	3	17
11	11	2008-03-20	2,008	3	20
12	12	2008-06-28	2,008	6	28
13	13	2008-07-12	2,008	7	12
14	14	2008-07-14	2,008	7	14
15	15	2008-08-01	2,008	8	1
16	16	2008-08-03	2,008	8	3
17	17	2008-09-04	2,008	9	4
18	18	2008-10-01	2,008	10	1
19	19	2008-10-08	2,008	10	8
20	20	2008-10-15	2,008	10	15
21	21	2008-10-28	2,008	10	28
22	22	2008-11-09	2,008	11	9
23	23	2008-11-10	2,008	11	10
24	24	2008-11-15	2,008	11	15

Fact_venta

SELECT * FROM Fact_venta

Enter a SQL expression to filter results (use Ctrl+Space)

	ID venta key	Product key	Fecha key	cantidad vendida	precio unidad	total venta
205	205	239	20	15	38	570
206	206	240	7	67	64	4,288
207	207	240	20	44	64	2,816
208	208	240	39	6	64	384
209	209	240	38	34	64	2,176
210	210	240	8	34	64	2,176
211	211	240	42	30	64	1,920
212	212	247	36	5	64	320
213	213	247	18	50	64	3,200
214	214	249	18	45	49	2,205
215	215	250	18	50	19	950
216	216	253	55	50	6	300
217	217	254	1	15	19	285
218	218	256	39	2	64	128
219	219	260	7	5	462	2,310
220	220	260	17	1	462	462
221	221	260	20	150	462	69,300
222	222	260	39	1	462	462
223	223	260	38	42	8	336
224	224	263	22	3	10	30
225	225	265	1	23	14	322
226	226	265	43	15	13	195