



UNIVERSIDAD DIGITAL DE ANTIOQUIA

Bases de Datos II

Periodo 2024-1

Profesor: VICTOR HUGO MERCADO

Alumno: Daniel Rojas Vidal

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

2024

1. Preparación

Revisión del Modelo Estrella:

Para cumplir con el primer paso de la preparación, revisaremos el modelo Estrella definido en el proyecto de análisis de la base de datos Jardinería, esto nos permitirá comprender la estructura y las relaciones entre las tablas de dimensiones y la tabla de hechos.

Verificación de la Base de Datos Staging:

El siguiente paso consiste en verificar la disponibilidad y consistencia de la base de datos de staging previamente creada en la actividad numero 2 así asegurarnos de que esta base de datos esté correctamente poblada y mantenga la misma estructura que la base de datos original es fundamental para el ETL.

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

Para realizar la extracción de datos desde la base de datos origen hacia la base de datos de Staging, utilizaremos consultas SQL para seleccionar y cargar los datos relevantes en las tablas correspondientes de la base de datos de Staging.

Ejemplo consultas de extracción:

```
INSERT INTO jardineria_IUD_staging.Categoria_producto SELECT *  
FROM jardineria_IUD.Categoria_producto;  
INSERT INTO jardineria_IUD_staging.cliente SELECT * FROM  
jardineria_IUD.cliente;  
INSERT INTO jardineria_IUD_staging.detalle_pedido SELECT * FROM  
jardineria_IUD.detalle_pedido;  
INSERT INTO jardineria_IUD_staging.empleado SELECT * FROM  
jardineria_IUD.empleado;  
INSERT INTO jardineria_IUD_staging.oficina SELECT * FROM  
jardineria_IUD.oficina;  
INSERT INTO jardineria_IUD_staging.pago SELECT * FROM  
jardineria_IUD.pago;  
INSERT INTO jardineria_IUD_staging.pedido SELECT * FROM  
jardineria_IUD.pedido;  
INSERT INTO jardineria_IUD_staging.producto SELECT * FROM  
jardineria_IUD.producto;
```

Verificación de la Integridad y Consistencia de los Datos:

Después de la extracción de datos, es importante verificar la integridad y consistencia de los datos extraídos para asegurar que cumplan con los requisitos del modelo Estrella. Esto puede incluir la verificación de la presencia de datos obligatorios, la validación de las relaciones entre las tablas y la identificación de posibles errores o inconsistencias.

3. Transformación de datos según las necesidades analíticas:

En esta etapa, aplicaremos técnicas de transformación de datos, como limpieza, normalización y enriquecimiento, para preparar los datos de acuerdo con las necesidades analíticas específicas.

La transformación de los datos se realizará utilizando consultas SQL

Ejemplo Limpieza

```
/* Limpieza de Datos (Eliminación de Registros Duplicados) */  
DELETE FROM jardineria_IUD_staging.cliente  
WHERE ID_cliente IN (  
    SELECT ID_cliente  
    FROM (  
        SELECT ID_cliente, ROW_NUMBER() OVER (PARTITION BY  
ID_cliente ORDER BY ID_cliente) AS rn  
        FROM jardineria_IUD_staging.cliente  
    ) t  
    WHERE rn > 1  
);
```

Ejemplo Normalización

```
/* Normalización de Datos (División de Campos de Texto) */  
UPDATE jardineria_IUD_staging.cliente  
SET nombre_cliente = LEFT(nombre_cliente, 50);
```

Ejemplo Enriquecimiento

```
/* Enriquecimiento de Datos (Agregar Nueva Columna Derivada) */  
ALTER TABLE jardineria_IUD_staging.cliente  
ADD COLUMN nombre_completo VARCHAR(100);  
/* Actualización de la nueva columna */  
UPDATE jardineria_IUD_staging.cliente  
SET nombre_completo = CONCAT(nombre_cliente, ' ',  
apellido_contacto);
```

4. Carga de registros en el Data Mart final:

En esta etapa, diseñaremos 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. Luego, ejecutaremos estas consultas de carga y verificaremos que los datos se hayan insertado correctamente en el Data Mart final.

Consultas SQL de Carga:

A continuación, se muestran ejemplos de consultas SQL para cargar registros transformados desde la base de datos de staging en el Data Mart final:

Ejemplo de carga en Data Mart

```
/* Carga de datos en las tablas del Data Mart */  
INSERT INTO DataMart.dim_producto (ID_producto, nombre,  
Categoria, dimensiones, proveedor, descripcion,  
cantidad_en_stock, precio_venta)  
SELECT ID_producto, nombre, Categoria, dimensiones, proveedor,  
descripcion, cantidad_en_stock, precio_venta  
FROM jardineria_IUD_staging.producto;
```

Verificación de los Datos Cargados:

Después de ejecutar las consultas de carga, es importante verificar que los datos se hayan insertado correctamente en el Data Mart final. Puedes realizar consultas de selección en las tablas del Data Mart final para verificar la integridad y consistencia de los datos cargados.

```
SELECT COUNT(*) FROM DataMart.dim_producto;  
SELECT COUNT(*) FROM DataMart.dim_cliente;  
SELECT COUNT(*) FROM DataMart.dim_empleado;  
SELECT COUNT(*) FROM DataMart.dim_tiempo;  
SELECT COUNT(*) FROM DataMart.hechos_ventas;
```

Adjunto el Script en el siguiente enlace, donde muestra el paso a paso realizado

[Ver script](#)