

Evidencia de aprendizaje 2. Creación de una base de datos de staging

Bases de Datos II

Ricardo García Duque

Edison Fabián Molina Herrera

Yamith Alejandro Castaño Pérez

Profesor:

Victor Hugo Mercado Ramos

Grupo.

PREICA2401B010094

Institución Universitaria Digital de Antioquia

2024.

	1
Introducción	2
Objetivos	3
Planteamiento del problema	4
Análisis del problema	5
Creación de la base de datos Staging	6
Conclusiones	9
Bibliografía	10

Introducción.

En la actualidad, el manejo acertado y preciso de la información es la herramienta imprescindible para lograr el éxito en las organizaciones, ordenar, almacenar, buscar e interpretar la información que se genera en las transacciones de la actividad económica es un mecanismo que asegura la comprensión y por ende el mejoramiento en las dinámicas productivas de nuestro negocio, este proceso es esencial, sin embargo, con la actualización constante de la tecnología y el procesamiento de datos, tenemos a mano mecanismos que nos permiten enfocar y organizar los datos de tal manera que se facilite el procesamiento de los mismos, podemos redirigir los enfoques hacia productos, servicios, métodos de venta, publicidad y ofertas de una manera novedosa y efectiva.

El uso de bases de datos de staging o bases de datos “preliminares”, permite al usuario perfilar los datos en un nivel de importancia o con un enfoque específico de acuerdo al uso que se desee y la información que se pretenda obtener del análisis, esto permite delimitar la visión y poder detectar información relevante de acuerdo al objetivo que se haya propuesto el usuario, proporcionando dimensiones acordes al contexto de la situación y abarcando los datos que se crean relevantes para la situación y desechando otros que por el contrario no añaden más relevancia a dichos datos.

Objetivos.

- Conocer el uso de herramientas para el manejo de bases de datos tales como sql server y Visual Studio community..
- Distinguir de una manera clara y lógica la información relevante que se desea incluir en la creación de la base de datos de staging para posteriormente usarla en la conformación de la dimensión de datos y consolidación en data-mart o data-warehouse.
- Comprender los diferentes mecanismos para establecer la conexión entre las bases de datos a usar y su respectiva configuración con respecto a las llaves primarias y relaciones.
- Emplear los conocimientos adquiridos para obtener una base de datos de staging funcional, ordenada y enfocada en la prioridad establecida por el usuario.

Planteamiento del problema.

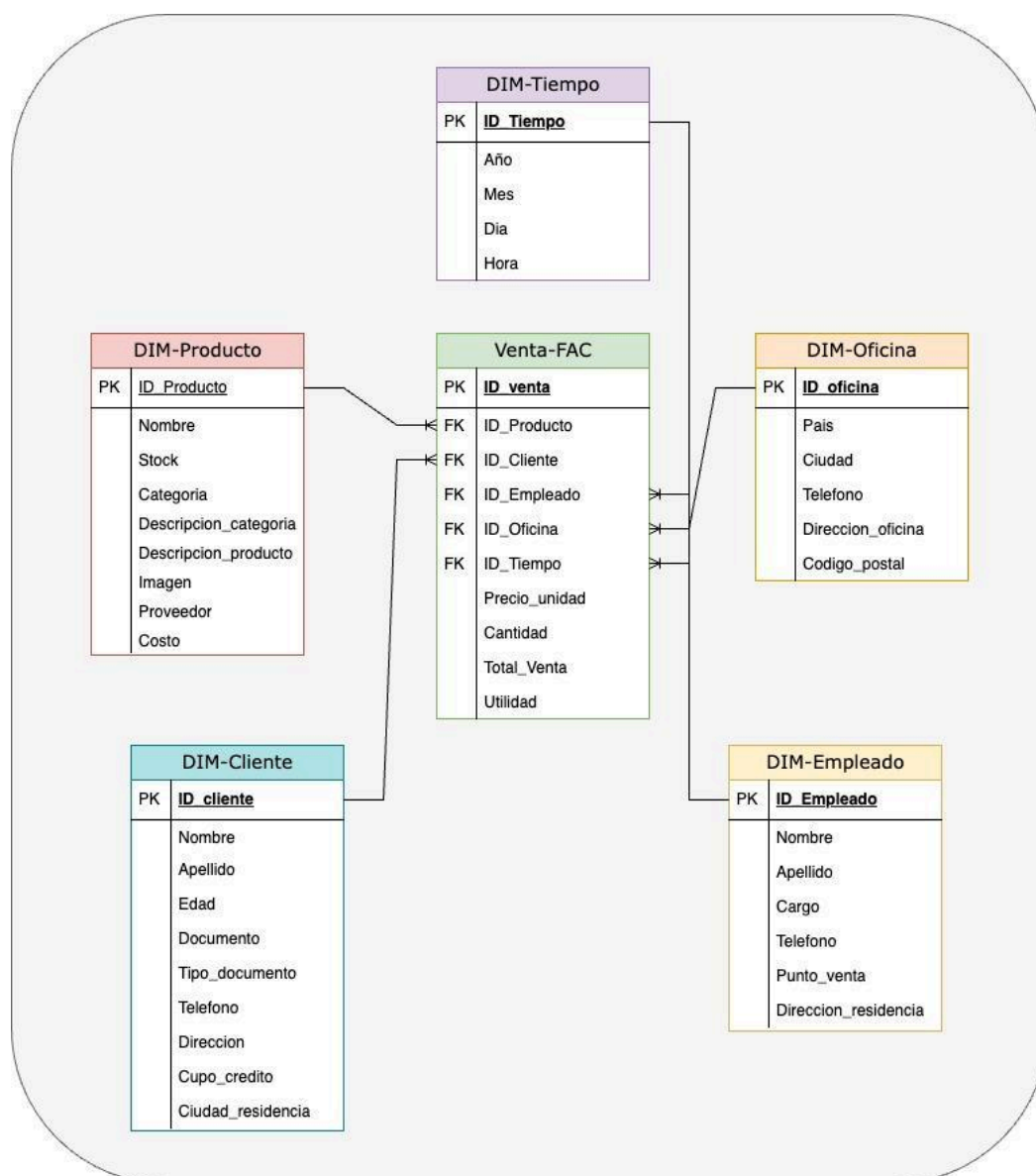
Se tiene la base de datos de Jardinería, la cual es una compilación de datos referentes a la actividad comercial de una distribuidora de productos agrícolas, se tiene stock de diversos productos como enseres, árboles, herramientas, abonos, flores, etc., así como los datos de los clientes, vendedores, ubicación geográfica, etc.

Debemos extraer campos de las tablas los cuales deben ser coherentes a la visión que se propone, la cual está orientada a construir una base de datos de staging basada en dimensiones construidas con campos de la mencionada base de datos los cuales deben tener características relevantes y deben alimentar las dimensiones propuestas, las cuales son producto, cliente, oficina, empleado, tiempo y factura.

Para lograr el cometido, se deben evaluar los campos de la base de datos jardinería con criterios claros y definidos previamente los cuales permitirán crear tablas completas y cuya información será relevante en el desarrollo de la base de datos de staging, la cual servirá como base para la elaboración de posteriores arreglos de datos más complejos.

Análisis del problema.

Inicialmente consideraremos la base de datos Jardinería y tomaremos como base para su análisis y procesamiento el modelo estrella elaborado a partir de dicha base de datos:



La información en el modelo estrella es pieza fundamental ya que nos brinda un escenario de dimensiones ya estructuradas y funcionales que nos sirven de base para la creación de la base de datos staging.

Creación de la base de datos Staging

A continuación se describe el paso a paso para la creación de la base de datos Staging, se realiza el siguiente proceso en el cual se mencionan las tablas y campos que aportan valor y ayudan con el objetivo para el análisis de la información:

Base de datos: Jardinería

Tabla: producto, categoria_producto:

```
SELECT
  ID_producto,nombre,Categoria,descripcion, proveedor, precio_venta,cantidad_en_stock,Desc_Categoria,dimensiones
FROM producto,Categoria_producto
```

unimos las dos tablas en una única tabla producto que abarca los campos importantes y segrega algunos poco relevantes o con la categoría null o vacío.

Tabla: pedido, detalle_pedido:

```
select pedido.ID_pedido,pedido.fecha_pedido,pedido.fecha_entrega,detalle_pedido.ID_producto,detalle_pedido.cantidad,
detalle_pedido.precio_unidad,
pedido.estado,pedido.ID_cliente
from pedido
Join detalle_pedido On pedido.ID_pedido=detalle_pedido.ID_pedido
```

unimos las dos tablas de pedidos para relacionar y emplear la información relevante, consideramos importante usar el tiempo ya que concierne más al hecho de saber el día que se realizó el pedido y el día que se entregó, este dato es indispensable para conocer el “ciclo del pedido”, y también el campo estado, el cual dictamina si se concluyó o no el proceso de compra.

Tabla: empleado, oficina:

```
select empleado.ID_empleado, empleado.nombre, empleado.apellido1, empleado.email, empleado.puesto,  
oficina.ID_oficina, oficina.telefono, oficina.Descripcion  
from empleado, oficina  
order by ID_empleado asc
```

Unimos los campos de empleado y oficina para crear una tabla complementaria en la información del empleado, así desecharmos campos irrelevantes como el segundo apellido y podremos ubicarlo geográficamente de manera adecuada.

Tabla: pago

```
select distinct pago.ID_pago, pago.ID_cliente, cliente.nombre_cliente, cliente.limite_credito,  
pago.fecha_pago, pago.id_transaccion, pago.total  
from pago, cliente  
order by pago.ID_cliente asc
```

Usamos la mayoría de campos de la tabla pagos ya que en su total eran muy relevantes, sumamos a esta tabla la información del cliente y su capacidad crediticia para crear una dimensión financiera lógica que involucre el pago y la capacidad adquisitiva del cliente.

Tabla: geográfica

```
select cliente.ID_cliente, cliente.pais, cliente.region, cliente.ciudad,  
cliente.codigo_postal, cliente.linea_direccion1, cliente.ID_empleado_rep_ventas,  
empleado.ID_oficina, oficina.Descripcion  
from empleado, cliente, oficina  
where cliente.ID_empleado_rep_ventas = empleado.ID_empleado  
order by cliente.pais
```

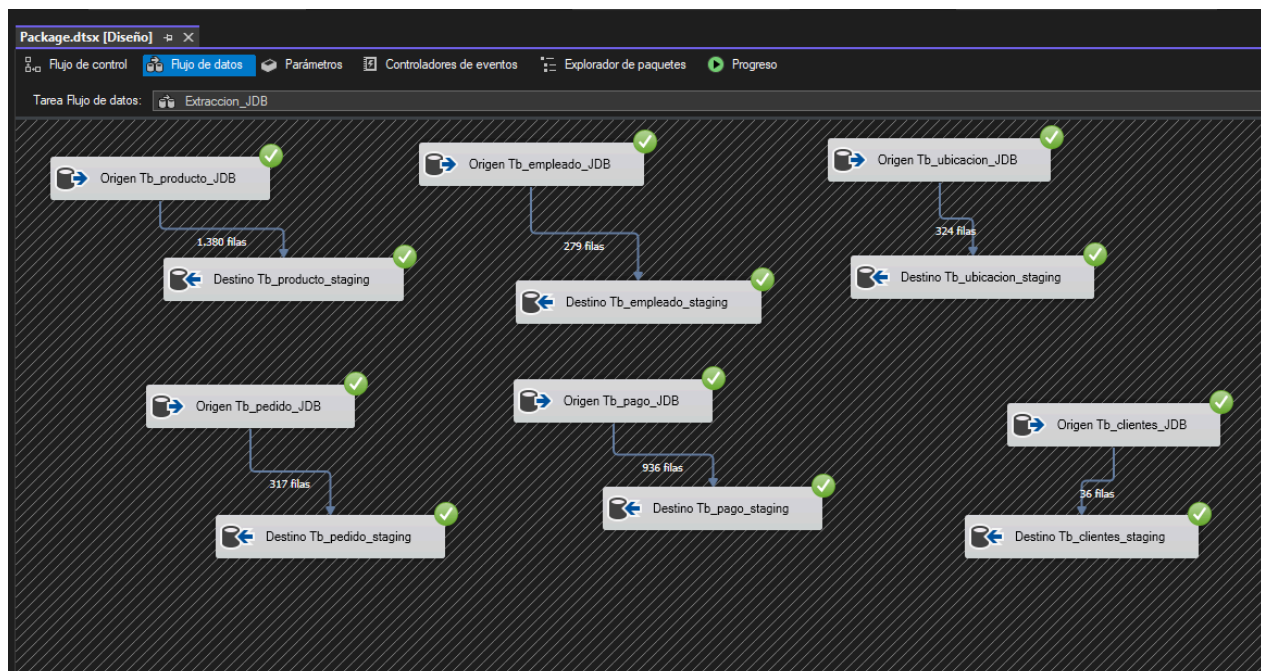
para la dimensión geográfica empleamos los datos relacionados al cliente en cuanto a su ubicación y a través del dato suministrado por el cliente del vendedor que realizó la venta, allí comparamos este id con el id del vendedor para extraer los datos correspondientes a la ubicación del vendedor y así relacionarlos geográficamente.

Tabla: clientes

```
select distinct cliente.ID_cliente, cliente.nombre_cliente,
cliente.telefono, cliente.limite_credito
from cliente
order by cliente.ID_cliente asc
```

Finalmente traemos los datos relevantes de la tabla cliente, ignoramos los datos de contacto y obviamente los de la ubicación ya que están considerados en la tabla anterior, adicionalmente consideramos el campo de límite de crédito el cual puede ser importante a la hora de discriminar clientes por capacidad y poder adquisitivo.

Como resultado obtenemos lo siguiente:



Conclusiones.

- En el desarrollo de esta actividad adquirimos conocimientos relevantes sobre el uso de herramientas para el manejo de bases de datos como SQL server y Visual Studio Community, que son fundamentales para poder diseñar y administrar información de una manera ágil y eficiente.
- Para la creación de la bases de datos staging, se tomó como base principal el modelo estrella realizado previamente, el cual nos ayudó a distinguir de manera clara y lógica la información relevante, esto nos permitió incluir datos específicos que nos ayudan a garantizar que vamos a contar con la información necesaria para el siguiente paso del proceso.
- Crear una base de datos staging es muy importante para crear tablas con campos más específicos, información más indispensable y con un enfoque más contextualizado.
- Al emplear estas herramientas de elaboración de flujos de datos y orígenes y destinos en el visual studio, aprendemos conceptos nuevos que son de utilidad para aplicaciones futuras.
- Finalmente con la aplicación de los conocimientos adquiridos en los módulos, investigación y en la socialización de los ejercicios desarrollados en clase pudimos desarrollar una base de datos staging funcional y ordenada, consolidando lo aprendido y el entendimiento apropiado que nos llevó a la solución.

Bibliografía

Rgward. (s. f.). *Introducción a SQL Server 2022 - Training*. Microsoft Learn.

<https://learn.microsoft.com/es-MX/training/modules/introduction-to-sql-server-2022/>

Staging: la salvaguarda de los procesos ETL. (s. f.).

<https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/312643/Staging-la-salvaguarda-de-los-procesos-ETL>