

S30 - Evidencia de aprendizaje 3

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

VIRGILIO RAMIREZ

ANGELO ANDRES DUQUE GARCIA

PATRICIA FRANCO RUIZ

INSTITUCIÓN UNIVERSITARIA DIGITAL DE ANTIOQUIA

Ingeniería en Desarrollo de Software

Bases de Datos II

VICTOR HUGO MERCADO - PREICA2401B010094

24 de mayo de 2024

Tabla de contenido

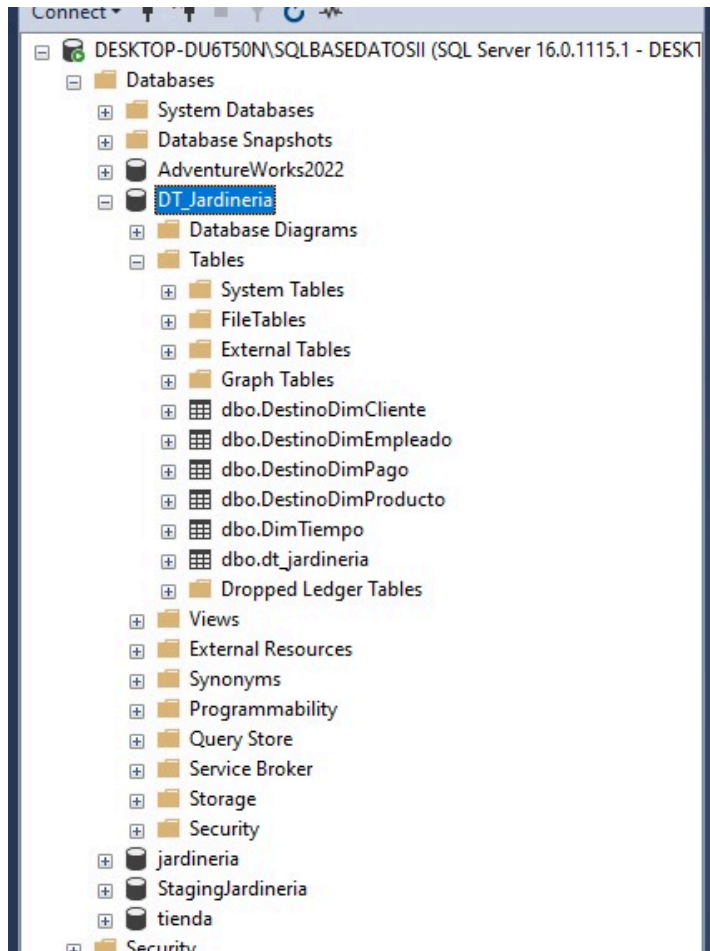
Tabla de contenido.....	2
Documentación y presentación:.....	3
• Carga de registros en el DataMart:.....	4
• Consulta SQL para extraer las fechas de pedidos únicas ordenadas por fecha:.....	5
• Consulta SQL para extraer información de los empleados:.....	6
• Consulta SQL para extraer información de los productos:.....	6
• Consulta SQL para extraer información de los clientes:.....	7
• Consulta SQL para extraer detalles de pedidos:.....	7
• Consulta SQL para extraer información de los pedidos:.....	8
• Consultas SQL para truncar (vaciar) varias tablas:.....	8
• Conexión a bases de datos origen y destino (Administración ADO.NET):.....	9
• Flujo de actividades:.....	10
• Flujo de control:.....	10
• Flujo de datos de producto:.....	11
• Salida del flujo:.....	11
Informe:.....	12
Objetivo del Informe:.....	12
Revisión del Modelo Estrella:.....	12
Verificación de la Base de Datos de Staging:.....	13
Conclusión.....	13

Documentación y presentación:

En primer lugar realizamos la revisión del modelo Estrella utilizado en el proyecto de análisis de la base de datos de Jardinería para entender cómo están organizadas y conectadas las tablas de dimensiones y la tabla de hechos.

Además, se comprobó que la base de datos de staging creada anteriormente para el proyecto estaba disponible y era consistente.

- **Carga de registros en el DataMart:**



La primera etapa del proceso ETL implica la extracción de datos desde la base de datos original de Jardinería. Para ello, se utilizaron varias consultas SQL para seleccionar y filtrar los datos necesarios de las tablas fuente.

- **Consulta SQL para extraer las fechas de pedidos únicas ordenadas por fecha:**

Especifique un administrador de conexiones de ADO.NET, un origen de datos o una vista del origen de datos, y seleccione el modo de acceso a los datos. Después de seleccionar el modo de acceso a los datos, seleccione entre las opciones adicionales de acceso a los datos que aparecerán.

Administrador de conexiones de ADO.NET:
origen jardineria

Modo de acceso a datos:
Comando SQL

Texto de comando SQL:
SELECT distinct(fecha_pedido)
FROM [jardineria].[dbo].[pedido]
ORDER BY fecha_pedido

Generar código
Examinar

- **Consulta SQL para extraer información de los empleados:**

Administrador de conexiones de ADO.NET:

origen jardineria

Modo de acceso a datos:

Comando SQL

Texto de comando SQL:

```
SELECT
ID_empleado,
nombre,
apellido1,
apellido2,
extension,
email,
ID_oficina,
ID_jefe,
puesto
FROM empleado;
```

Ge

- **Consulta SQL para extraer información de los productos:**

Administrador de conexiones de ADO.NET:

origen jardineria

Modo de acceso a datos:

Comando SQL

Texto de comando SQL:

```
SELECT
ID_producto,
CodigoProducto,
nombre,
Categoria,
dimensiones,
proveedor,
descripcion,
cantidad_en_stock,
precio_venta,
precio_proveedor
FROM producto;
```

G

- **Consulta SQL para extraer información de los clientes:**

Administrador de conexiones de ADO.NET:

origen jardineria

Modo de acceso a datos:

Comando SQL

Texto de comando SQL:

```
SELECT
  ID_cliente,
  nombre_cliente,
  nombre_contacto,
  apellido_contacto,
  telefono,
  fax,
  linea_direccion1,
  linea_direccion2,
  ciudad,
  region,
  pais,
  codigo_postal,
  ID_empleado_rep_ventas,
  limite_credito
FROM cliente;
```

- **Consulta SQL para extraer detalles de pedidos:**

Especifique un administrador de conexiones de ADO.NET, un origen de datos o una vista del origen de datos, y seleccione el modo de acceso a los datos. Después de seleccionar el modo de acceso a los datos, seleccione entre las opciones adicionales de acceso a los datos que aparecerán.

Administrador de conexiones de ADO.NET:

origen jardineria

Nueva...

Modo de acceso a datos:

Comando SQL

Texto de comando SQL:

```
SELECT
  ID_pedido,
  ID_producto,
  cantidad,
  precio_unidad,
  numero_linea
FROM detalle_pedido;
```

Generar consulta...

Examinar...

- Consulta SQL para extraer información de los pedidos:

Administrador de conexiones de ADO.NET:

origen jardineria

Modo de acceso a datos:

Comando SQL

Texto de comando SQL:

```
SELECT
  ID_pedido,
  fecha_pedido,
  fecha_esperada,
  fecha_entrega,
  estado,
  comentarios,
  ID_cliente
FROM pedido;
```

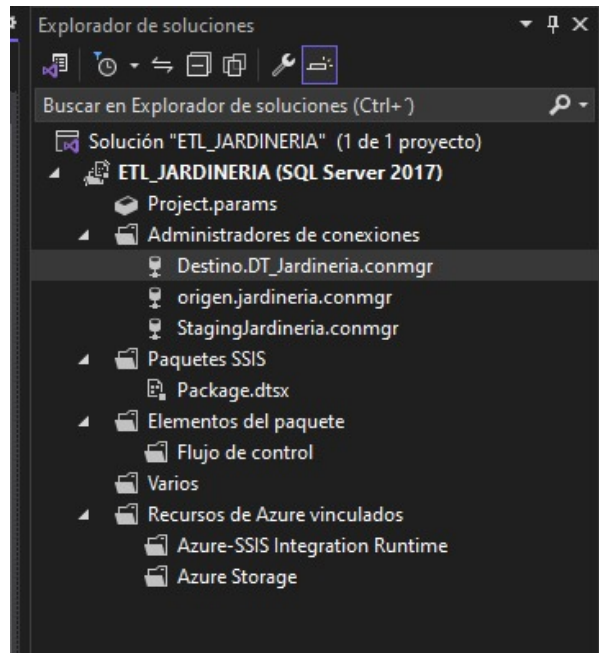
- Consultas SQL para truncar (vaciar) varias tablas:

Escribir consulta SQL

```
TRUNCATE TABLE [dbo].[Destino_TB_Cliente]
TRUNCATE TABLE [dbo].[Destino_TB_DetallePedido]
TRUNCATE TABLE [dbo].[Destino_TB_DetalleTiempo]
TRUNCATE TABLE [dbo].[Destino_TB_Empleado]
TRUNCATE TABLE [dbo].[Destino_TB_Producto]
TRUNCATE TABLE [dbo].[stating_Pedido]
```

Aceptar Cancelar

- **Conexión a bases de datos origen y destino (Administración ADO.NET):**



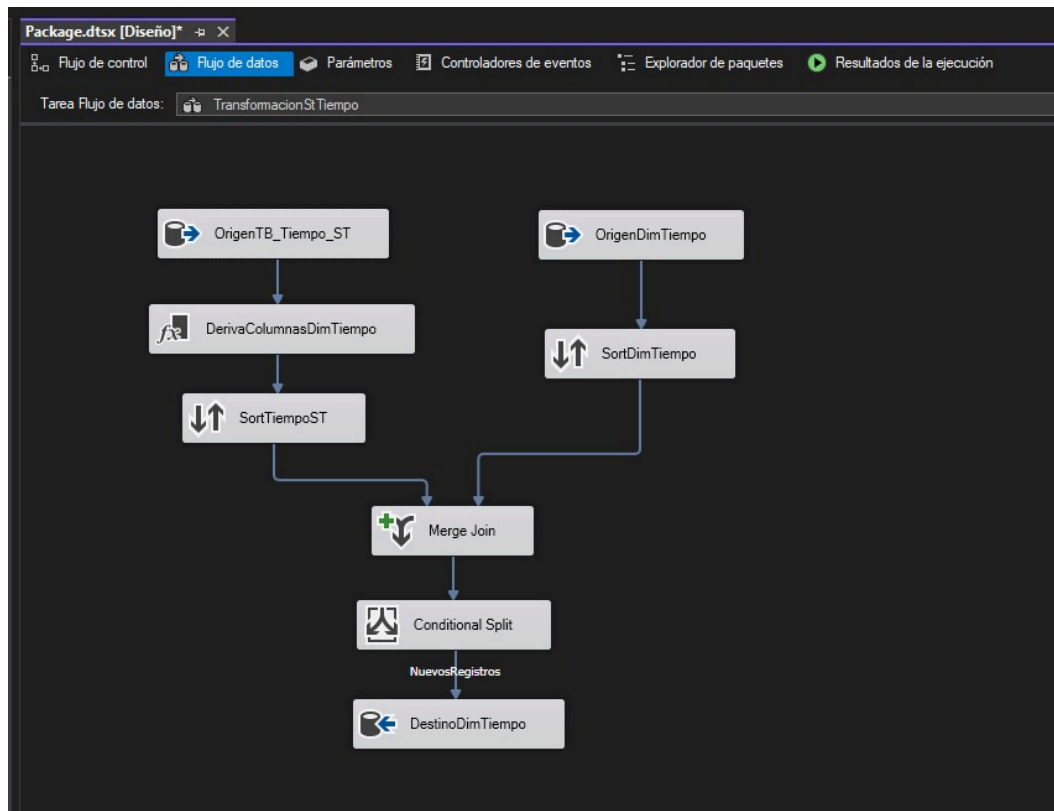
- Flujo de actividades:



- Flujo de control:



- Flujo de datos de producto:



- Salida del flujo:

Salida

Mostrar salida de: Depurar

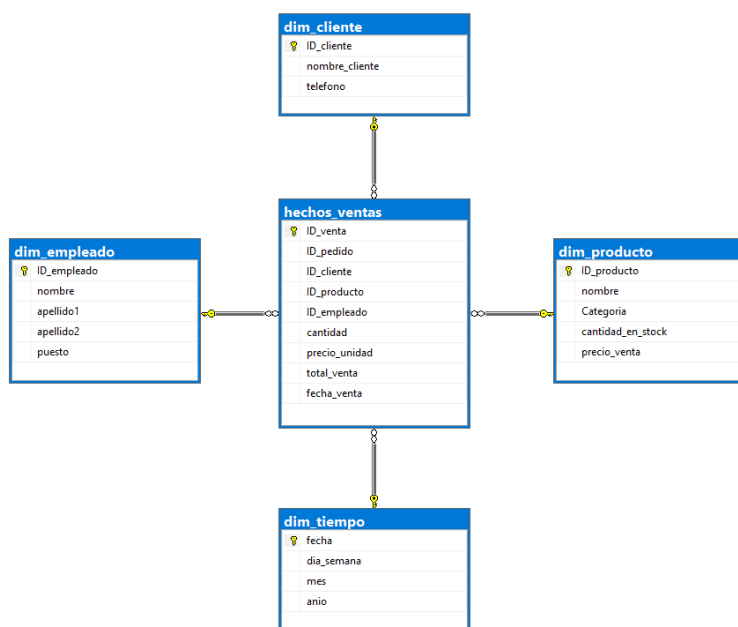
Información: 0x40043008 en TransformacionStPago, SSIS.Pipeline: "DestinoDimPago" wrote 0 rows.
Información: 0x40043009 en TransformacionStPago, SSIS.Pipeline: Cleanup phase is beginning.
Información: 0x4004300A en TransformacionStProducto, SSIS.Pipeline: Validation phase is beginning.
Advertencia: 0x80047076 en TransformacionStProducto, SSIS.Pipeline: The output column "Categoria" (65) on output "Merge Join Output" (56) and component "Merge Join" (37) is not in the output schema.
Advertencia: 0x80047076 en TransformacionStProducto, SSIS.Pipeline: The output column "descripcion" (71) on output "Merge Join Output" (56) and component "Merge Join" (37) is not in the output schema.
Advertencia: 0x80047076 en TransformacionStProducto, SSIS.Pipeline: The output column "cantidad_en_stock" (73) on output "Merge Join Output" (56) and component "Merge Join" (37) is not in the output schema.
Advertencia: 0x80047076 en TransformacionStProducto, SSIS.Pipeline: The output column "precio_venta" (75) on output "Merge Join Output" (56) and component "Merge Join" (37) is not in the output schema.
Advertencia: 0x80047076 en TransformacionStProducto, SSIS.Pipeline: The output column "precio_proveedor" (77) on output "Merge Join Output" (56) and component "Merge Join" (37) is not in the output schema.
Advertencia: 0x80047076 en TransformacionStProducto, SSIS.Pipeline: The output column "ID_producto" (241) on output "Sort Output" (238) and component "SortDimProducto" (216) is not in the output schema.
Advertencia: 0x80047076 en TransformacionStProducto, SSIS.Pipeline: The output column "CodigoProducto" (243) on output "Sort Output" (238) and component "SortDimProducto" (216) is not in the output schema.
Advertencia: 0x80047076 en TransformacionStProducto, SSIS.Pipeline: The output column "nombre" (245) on output "Sort Output" (238) and component "SortDimProducto" (216) is not in the output schema.
Advertencia: 0x80047076 en TransformacionStProducto, SSIS.Pipeline: The output column "dimensiones" (247) on output "Sort Output" (238) and component "SortDimProducto" (216) is not in the output schema.
Advertencia: 0x80047076 en TransformacionStProducto, SSIS.Pipeline: The output column "proveedor" (249) on output "Sort Output" (238) and component "SortDimProducto" (216) is not in the output schema.
Información: 0x4004300B en TransformacionStProducto, SSIS.Pipeline: Prepare for Execute phase is beginning.
Información: 0x4004300C en TransformacionStProducto, SSIS.Pipeline: Pre-Execute phase is beginning.
Información: 0x4004300D en TransformacionStProducto, SSIS.Pipeline: Execute phase is beginning.
Información: 0x4004300E en TransformacionStProducto, SSIS.Pipeline: Post Execute phase is beginning.
Información: 0x4004300F en TransformacionStProducto, SSIS.Pipeline: "DestinoDimProducto" wrote 0 rows.
Información: 0x40043010 en TransformacionStProducto, SSIS.Pipeline: Cleanup phase is beginning.
Paquete SSIS "C:\Users\FAMILIA\source\repos\ETL JARDINERIA\ETL JARDINERIA\Package.dtsx" finalizado: Correcto.

Informe:

Objetivo del Informe:

En este documento, detallamos el proceso completo de transformación de datos (ETL - Extracción, Transformación y Carga) utilizado para poblar el Data Mart final, en el contexto del proyecto de análisis de la base de datos de Jardinería. Este proceso se estructura en varias etapas, cada una de las cuales se describe a continuación, junto con las consultas SQL empleadas y las técnicas de ETL aplicadas. Además, se hará referencia al modelo Estrella utilizado en este proyecto.

Revisión del Modelo Estrella:



El modelo Estrella definido en el proyecto de análisis de la base de datos Jardinería se compone de las siguientes tablas:

Tabla de Hechos:	Ventas
Tablas de Dimensiones:	Producto, Cliente, Tiempo, Empleado

De esta manera se confirma que el modelo estrella nos permite realizar un análisis de ventas a través de múltiples dimensiones, permitiendo consultas rápidas y eficientes.

Verificación de la Base de Datos de Staging:

Se revisó la disponibilidad y consistencia de la base de datos de staging, asegurando que las tablas de staging estén preparadas para recibir los datos extraídos de la base de datos origen.

Conclusión

El proceso ETL garantiza que los datos estén listos para el análisis empresarial, lo que facilita el análisis del producto más vendido y otros puntos de decisión importantes. Mientras desarrollamos esta actividad, nos dimos cuenta de lo importante que es realizar el proceso correctamente para obtener información clara, integral y óptima.

Cuando cada paso del proceso se describe clara y completamente, podemos lograr mejor los objetivos propuestos, lo que garantiza la calidad y coherencia de los datos finales del Data

Mart. Esto no sólo garantiza la precisión y confiabilidad de los datos, sino que también proporciona una base sólida para el análisis y la generación de informes.

El uso de un proceso ETL bien estructurado y documentado proporciona varios beneficios importantes:

Calidad de los datos: la limpieza y transformación de los datos garantiza que los datos estén libres de errores, duplicados y valores inconsistentes.

Coherencia: los datos se normalizan y estructuran de manera consistente, lo que facilita su uso en análisis posteriores.

Facilidad de análisis: los datos organizados en un modelo en estrella facilitan la formulación de preguntas complejas y la creación de informes detallados.

Mejor toma de decisiones: la información precisa y actualizada mejora la capacidad de tomar decisiones informadas y estratégicas.

Logro de objetivos: documentar cada paso del proceso permite un seguimiento claro del progreso y garantiza el logro efectivo de los objetivos del proyecto.

En conclusión, el proceso ETL es fundamental no solo para preparar los datos para el análisis empresarial, sino también para mantener la integridad y utilidad a largo plazo de los datos. La implementación y documentación cuidadosas de cada paso del proceso es esencial para garantizar que el Data Mart final cumpla con los estándares de calidad y sirva como una herramienta eficaz de análisis y toma de decisiones.