

Actividad: EA2_Grupo#CreacionBDStaging_Jardineria

Asignatura: Bases de Datos II

PREICA2401B010093

PREICA2401B010076

Docente: Víctor Hugo Mercado

Estudiante:

Elizabeth Alzate

Jimmy Luis Mora Russy

Vanessa Foronda M.

Ingeniería Software y Datos

Institución: IU Digital de Antioquia

Fecha: 05 mayo 2024



OBJETIVO

Construir consultas SQL para la creación de una base de datos staging basada en el modelo relacional de la base de datos *Jardinería*. Consúltala aquí.

1. Análisis de los datos almacenados en la base de datos *Jardinería*:
 - Revisar los datos almacenados en el *Jardinería* para identificar cuáles son relevantes y cuáles se deben trasladar a la base de datos *Staging*.
2. Construcción de la base de Datos *Staging*:
 - Diseñar la estructura de tablas que estarán en la base de datos *Staging*.
 - Construir las consultas que permitan traer los registros de *Jardinería* a la Base de Datos *Staging*.
 - Ejecutar las consultas y validar que los datos queden almacenados correctamente en la Base de datos *Staging*.
 - Construir el BK de ambas bases de datos.
3. Documentación y presentación:

Redactar un documento escrito que cumpla con normas APA y contenga:

 - Introducción
 - Objetivos
 - Planteamiento del problema
 - Análisis del problema
 - Propuesta de la solución con:
 - Correcciones a la entrega 1.
 - Descripción del análisis realizado a los datos *Jardinería* y cómo estos se trasladaron a la base de datos *Staging*.
 - Anexos: debe agregar los BK de las dos bases de datos y el documento del *script* de las consultas para crear la base de datos *Staging*.
 - Bibliografía.

ÍNDICE: BASE DE DATOS STAGING

1. Introducción	3
2. Objetivos	4
3. Planteamiento del problema	4
4. Análisis del problema	5
5. Propuesta de la solución	6
a. Correcciones a la entrega 1	6
b. Descripción del análisis realizado a los datos Jardinería y cómo estos se trasladaron a la base de datos Staging.	7
c. Anexos	8
6. Bibliografía	8

DESARROLLO

1. Introducción

El presente trabajo se ha realizado con el fin de crear un staging como parte de la arquitectura del datawarehouse (DWH), requerido para estructurar la información de la empresa de Jardinería, la cual cuenta con datos distribuidos en varias tablas en una base de datos, y que se ha diagramado y estudiado en la actividad No.1 con el modelo estrella.

Para esta actividad, se han corregido las novedades del diagrama de la Actividad No.1, y se ha estructurado el Staging acorde con el análisis del equipo de trabajo, para el modelo estrella del proceso de ventas de la empresa Jardinería.

Se considera que este staging facilitará la extracción de los datos, mejorará la calidad de estos y permitirá acceder en detalle a la información contenida en el DWH, que se llegará a implementar en el caso ideal, en caso contrario a no alcanzar el ideal, este staging facilitará el proceso de transformación de los datos y podrá ser utilizado en herramientas de visualización de business inteligencia para el procesamiento y toma de decisiones.

2. Objetivos

Construir consultas SQL para la creación de una base de datos staging basada en el modelo relacional de la base de datos Jardinería, de tal manera que se permita un análisis de los datos almacenados, la construcción de la base de datos staging, el diseño de la estructura de tablas destino, la construcción que permite traer los registros de la tabla Origen a la Destino, la ejecución de las consultas y validación de los datos, y la construcción del BK de ambas.

3. Planteamiento del problema

La empresa Jardinería tiene mucha información y datos en diferentes fuentes de esta, sin embargo, no contiene una herramienta que le permita reunir estos datos y garantizar a su vez que los datos sean consistentes con la realidad, sean válidos y estructurados. Siendo así, han contratado a un equipo experto en la consolidación, almacenamiento y procesamiento de datos para inteligencia de negocios, con el fin de estructurar su data mart en el proceso de ventas. Este equipo les ayudará a integrar los datos de tal manera que queden limpios y consistentes, proporcionará un entorno de validación y pruebas antes de continuar al entorno de producción (una base de datos staging), optimizando el

rendimiento de las consultas y procesos relacionados. La consolidación de datos, la seguridad y el control de acceso a estos datos deberá acompañar el proceso.

En resumen, se podría decir que el problema radica en que Jardinería tiene muchas diferentes fuentes de información y datos, no tiene estructura ni conocimiento de estos datos, conocer la situación actual, qué información brindan los datos que no se han tenido en cuenta, lo que le dificulta crear estrategias de mejora, ya que su procesamiento de datos es lento, poco confiable e inseguro.

4. Análisis del problema

La empresa Jardinería desea crear una base Staging para el departamento comercial, que le permita tener información compleja de las transacciones de sus pedidos. Actualmente la empresa no cuenta con modelos estructurados de información que les permita ser confiables en el desarrollo de estrategias por falta de un pobre conocimiento en sus datos claves. Siendo así, se vuelve necesario contratar al equipo de expertos en análisis y procesamiento de datos, que realicen un staging de sus fuentes de datos por medio de las siguientes actividades entre otras:

Limpieza de datos: Asegurará la integridad y consistencia de los datos en la base staging limpiando y transformando los datos (antes de cargarlos en la base final -DWH-).

Validación de datos: Las pruebas en un entorno controlado que no afecte la fuente de datos inicial ni los datos en el entorno de producción, serán perfectos para asegurar la calidad y la validación de los datos.

Pruebas de rendimiento: En el staging también es posible realizar las consultas antes de implementar los cambios y transferirlos al entorno de producción, lo que permitirá rendimiento y optimización.

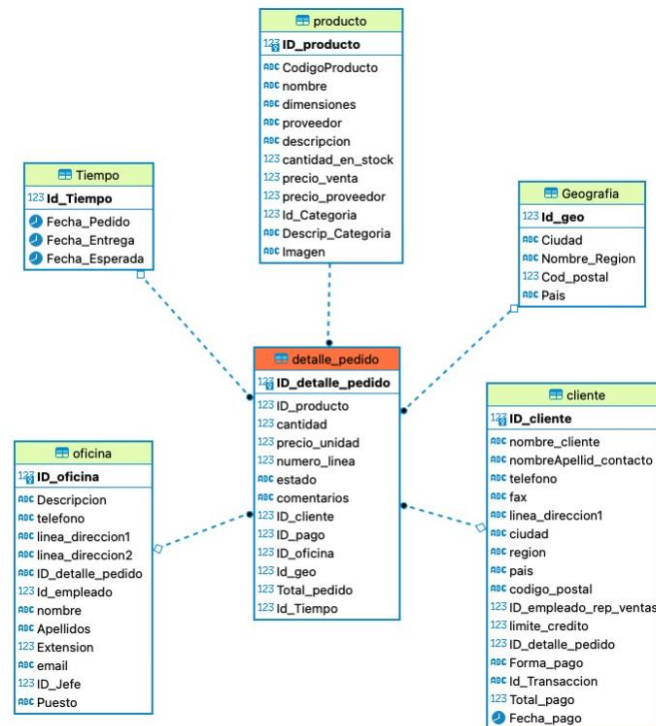
Consolidación de datos: Con el staging (la base de preparación), es posible recopilar de diferentes fuentes de información en diferentes formatos (Operacionales, ERP, CRM, PLM, datos Semiestructurados / No estructurados) y consolidar en una sola base de datos (Staging).

Seguridad y control de la información: permite definir quienes serán los roles / perfiles o personas autorizadas que podrá acceder y manipular los datos antes de la carga en el DWH.

Actualmente, ninguna de esas actividades está contemplada en el procesamiento, custodia, análisis de los datos e información de la compañía, lo que podría resultar en pérdidas de datos, datos en manos malintencionadas, información errónea para la toma de decisiones, lo que afecta directamente los indicadores de la compañía.

5. Propuesta de la solución

a. Correcciones a la entrega 1



- Cambio: Se crea dimensión geografía con los campos:
 - Ciudad
 - Region
 - Cod_postal
 - Pais
- Cambio: Se crea dimensión Tiempo con los campos:
 - Fecha_Pedido,
 - Fecha_Entrega
 - Fecha_Esperada
- Cambio: Se unifica la dimensión pago con la dimensión clientes y se trasladan los campos:
 - id_detalle_pedido
 - forma_pago
 - id_transaccion
 - Total_pago
 - Fecha_pago
- Cambio: Se agrega columna total a la tabla de hechos detalle_pedido.

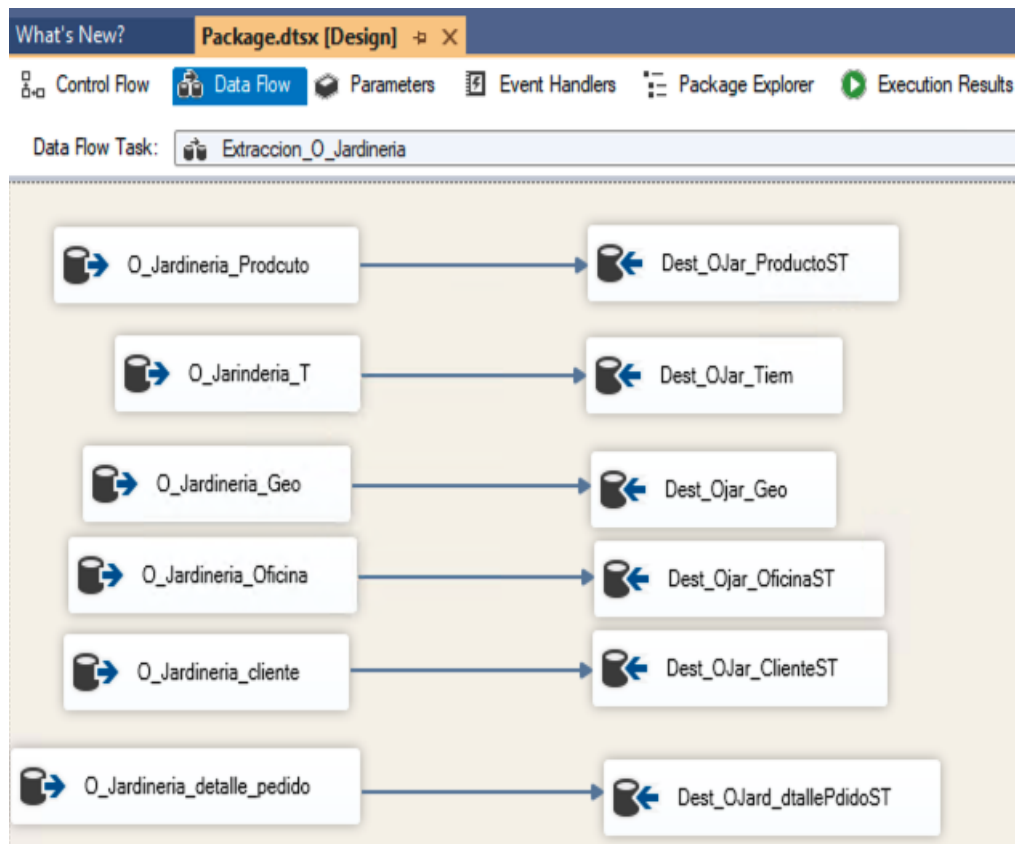
b. Descripción del análisis realizado a los datos *Jardinería* y cómo estos se trasladaron a la base de datos *Staging*.

Se analizaron las relaciones entre las dimensiones y la tabla de hechos en la base de datos jardinería, con el fin de establecer el orden de cada una de las tablas a trasladar a la base de datos Staging. Se buscan las relaciones de las tablas a través de sus PK y FK.

Posteriormente a esto, se inicia la extracción utilizando la herramienta Visual Studio 2022 creada por Microsoft. Al abrir esta, se realiza la conexión a las bases de datos origen (Jardinería) y destino (StagingJardineria). Una vez verificadas las conexiones en nuestro caso de tipo ADO.NET tanto en el origen como en destino, generamos una tarea de flujo de datos y sobre esta, en el flujo de datos realizamos la conexión entre cada tabla de origen y destino, y conectando a través de JOIN aquellas dimensiones unificadas en una sola, como parte del proceso de optimización. Por último, se dejó la tabla de hechos, en la cual se realizaría la conexión entre las dimensiones: Geografía, tiempo, producto, cliente y oficina. Puesta el enlace a esta se haría a través de las FK de las dimensiones asociadas a la tabla de hechos.

Luego de este proceso, se ejecuta el flujo de datos, obteniendo como resultado, el traslado de la información de las tablas de jardinería a las tablas de la base Staging.

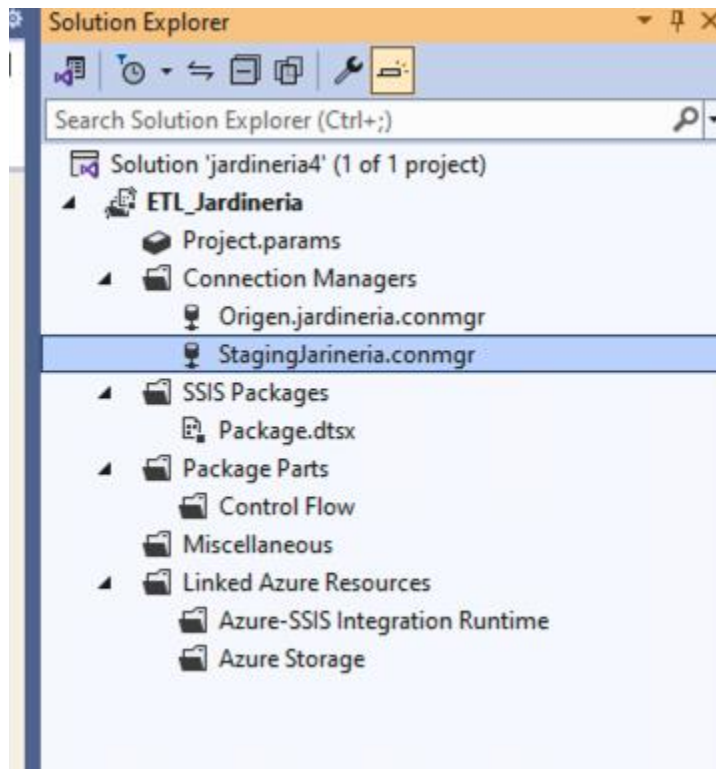
Creación data flow



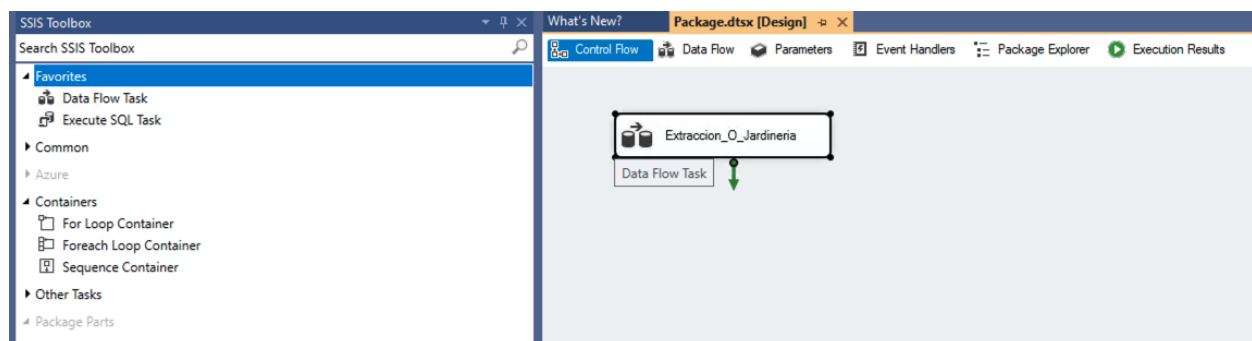
Ejecución del flujo de datos



BD StagingJardinería



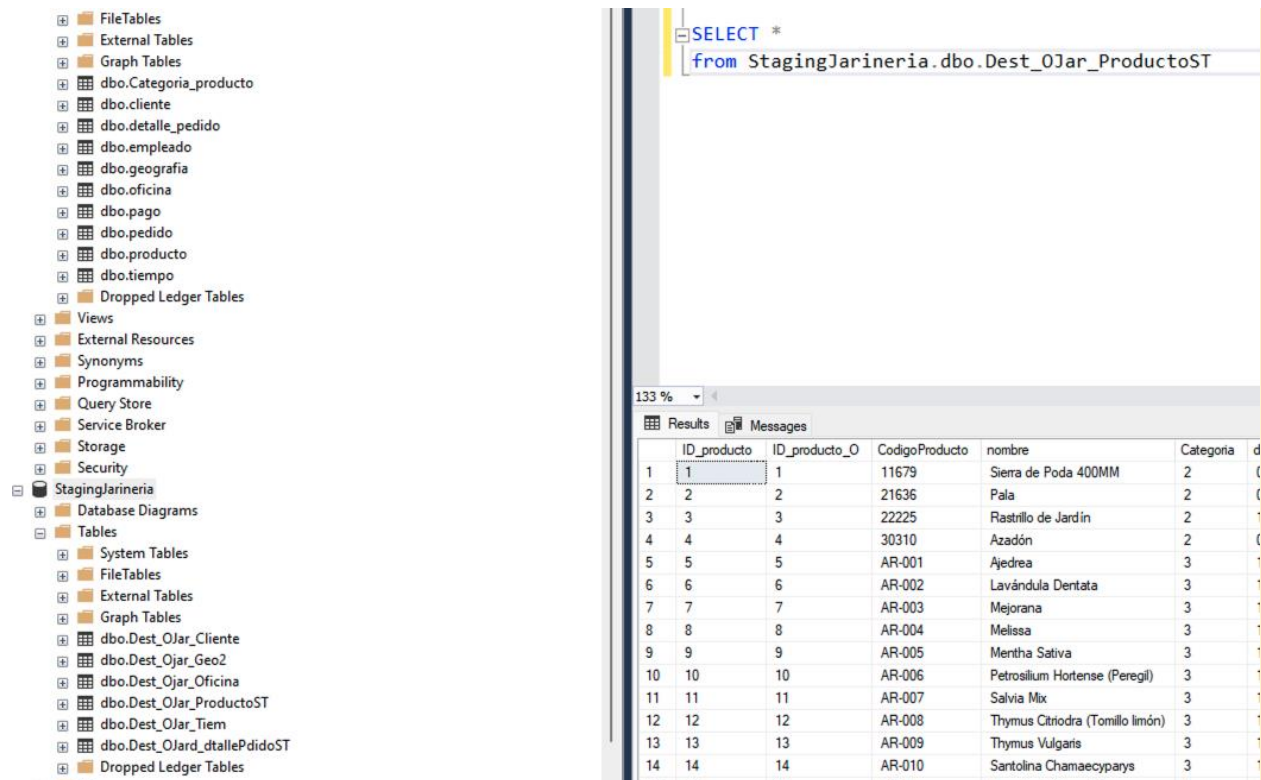
Extracción de datos



Creación de la base de datos StagingJardinería



Prueba de extracción de datos de la tabla producto de Jardinería a StagingJardineria.



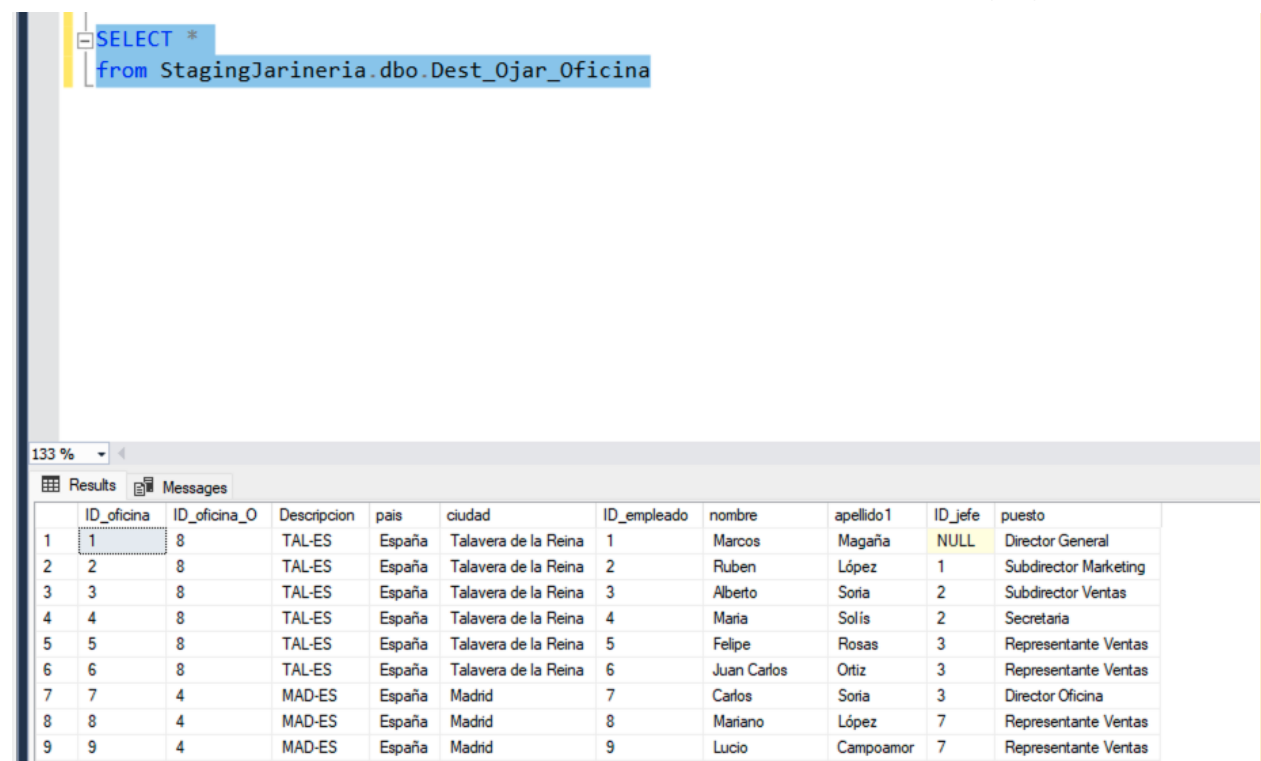
The screenshot shows the SQL Server Enterprise Manager interface. On the left, the 'StagingJardineria' database is selected, showing its tables. The 'dbo.dest_OJar_ProductoST' table is highlighted. On the right, a query window displays the following SQL statement:

```
SELECT *
from StagingJardineria.dbo.dest_OJar_ProductoST
```

Below the query window, the 'Results' tab shows the data extracted from the table. The results are as follows:

ID_producto	ID_producto_O	CodigoProducto	nombre	Categoria
1	1	11679	Sierra de Poda 400MM	2
2	2	21636	Pala	2
3	3	22225	Rastrillo de Jardín	2
4	4	30310	Azadón	2
5	5	AR-001	Ajedrea	3
6	6	AR-002	Lavándula Dentata	3
7	7	AR-003	Mejorana	3
8	8	AR-004	Melissa	3
9	9	AR-005	Mentha Sativa	3
10	10	AR-006	Petrosilium Hortense (Peregil)	3
11	11	AR-007	Salvia Mix	3
12	12	AR-008	Thymus Citriodra (Tomillo limón)	3
13	13	AR-009	Thymus Vulgaris	3
14	14	AR-010	Santolina Chamaecypariss	3

Prueba de extracción de datos de la tabla oficina de Jardinería a StagingJardineria



The screenshot shows the SQL Server Enterprise Manager interface. On the left, the 'StagingJardineria' database is selected, showing its tables. The 'dbo.dest_OJar_Oficina' table is highlighted. On the right, a query window displays the following SQL statement:

```
SELECT *
from StagingJardineria.dbo.dest_OJar_Oficina
```

Below the query window, the 'Results' tab shows the data extracted from the table. The results are as follows:

ID_oficina	ID_oficina_O	Descripcion	pais	ciudad	ID_empleado	nombre	apellido1	ID_jefe	puesto
1	8	TAL-ES	España	Talavera de la Reina	1	Marcos	Magaña	NULL	Director General
2	8	TAL-ES	España	Talavera de la Reina	2	Ruben	López	1	Subdirector Marketing
3	8	TAL-ES	España	Talavera de la Reina	3	Alberto	Soria	2	Subdirector Ventas
4	8	TAL-ES	España	Talavera de la Reina	4	Maria	Solis	2	Secretaria
5	8	TAL-ES	España	Talavera de la Reina	5	Felipe	Rosas	3	Representante Ventas
6	8	TAL-ES	España	Talavera de la Reina	6	Juan Carlos	Ortiz	3	Representante Ventas
7	4	MAD-ES	España	Madrid	7	Carlos	Soria	3	Director Oficina
8	4	MAD-ES	España	Madrid	8	Mariano	López	7	Representante Ventas
9	4	MAD-ES	España	Madrid	9	Lucio	Campoamor	7	Representante Ventas

- Análisis de los datos almacenados en la base de datos Jardinería:
- Revisar los datos almacenados en Jardinería para identificar cuáles son relevantes y cuáles se deben trasladar a la base de datos Staging.

Hemos analizado e identificado los datos almacenados en la base de datos Jardinería, a continuación, los que encontramos relevantes y cuales consideramos se deben trasladar a la Base de datos Staging:

Item	Tabla Evaluada	Dato Relevante	Se debe trasladar a Staging	Observaciones
1	Cliente	cliente.ID_cliente cliente.nombre_cliente cliente.nombre_contacto cliente.apellido_contacto cliente.ciudad cliente.pais cliente.codigo_postal cliente.ID_empleado_rep_ventas cliente.limite_credito pago.forma_pago pago.id_transaccion pago.fecha_pago, pago.total pago.ID_pago	Los datos de la columna relevante son los datos para trasladar a Staging	Algunos campos se extraen para a tabla geografía
2	Producto	ID_producto CodigoProducto cantidad_en_stock precio_venta precio_proveedor Categoria proveedor nombre	Los datos de la columna relevante son los datos para trasladar a Staging	
3	Tiempo	Id_Tiempo Fecha_Pedido Fecha_Entrega Fecha_Esperada	Los datos de la columna relevante son los datos para trasladar a Staging	

4	Geografía	Id_geo Ciudad, Cod_postal Pais	Los datos de la columna relevante son los datos para trasladar a Staging	Se extrae de la tabla oficina y cliente
5	Oficina	oficina.ID_oficina oficina.Descripcion oficina.codigo_postal oficina.pais oficina.ciudad empleado.ID_empleado empleado.nombre empleado.apellido1 empleado.ID_jefe empleado.puesto	Los datos de la columna relevante son los datos para trasladar a Staging	Algunos campos se extraen para a tabla geografía
6	Hechos: Detalle_Producto	detalle_pedido.ID_detalle_pedido detalle_pedido.ID_producto detalle_pedido.cantidad detalle_pedido.precio_unidad detalle_pedido.numero_linea detalle_pedido.estado detalle_pedido.comentarios cliente.ID_cliente detalle_pedido.ID_pago oficina.ID_oficina Geografia.Id_geo detalle_pedido.Total_pedido Tiempo.Id_Tiempo	Los datos de la columna relevante son los datos para trasladar a Staging	Tabla principal de hechos, contiene los datos propios de la tabla y las FK de las tablas relacionadas

c. Anexos

Enlace archivos: Base de datos staging, jardinería y proyecto en Visual Studio.
<https://drive.google.com/drive/folders/1U18VwoINPxieE4wyOLM6xyHyhDbV5k0->

Enlace scripts

<https://docs.google.com/document/d/1p3ZBFSAIIXcO0cJT-xsk9VcsS5IJCneW/edit>

Se agrega al presente documento:

BK de la base de datos Origen

BK de la base de datos Destino

Documento con script de las consultas para la creación de la BD Staging.

Mi unidad > 2024 Bases de Datos II ... > 2024_04_abril_15 ▾

Tipo ▾ Personas ▾ Modificado ▾

Nombre	Propietario	Última modifica...	Tamaño de a	
Creacion.Jardineria.zip	yo	15 abr 2024 yo	20 kB	
Staging.Jardineria.bak	yo	9 may 2024 yo	17,8 MB	
jardineria.bak	yo	9 may 2024 yo	16,6 MB	

6. Bibliografía

https://cursos.yura.website/wp-content/uploads/2020/03/Introduccion_al_Business_Intelligence.pdf

<https://revistas.udistrital.edu.co/index.php/vinculos/article/view/15592/15391>

https://openaccess.uoc.edu/bitstream/10609/136207/5/Fundamentos%20de%20inteligencia%20de%20negocio_Mo%C2%BFdulo2_Disen%C2%BFo%20de%20un%20data%20warehouse.pdf