

Evidencia de aprendizaje 4. Github y Pruebas

Bases de Datos II (PREICA2501B010095)

Indira Johanna Hamdam Jarava

Juan Sebastián Clavijo Correa

Cristian José Vicioso Escamilla

Docente

Víctor Hugo Mercado

Ingeniería de Software y Datos

Institución Universitaria Digital de Antioquia

2025

Introducción

En el ámbito de la gestión de bases de datos, la migración y optimización de estructuras de almacenamiento de información son procesos fundamentales para garantizar la eficiencia y calidad de los datos. En este trabajo, se aborda la creación de una base de datos de tipo Staging, cuyo objetivo principal es servir como un espacio intermedio para la migración y transformación de los datos provenientes de la base de datos Jardinería.

La base de datos Jardinería contiene información relacionada con clientes, productos, pedidos y detalles de pedidos, los cuales requieren una estructura organizada y optimizada para su posterior análisis. Mediante la implementación de una base Staging, se facilita el proceso de depuración, validación y estandarización de los datos antes de su integración en un sistema definitivo de almacenamiento o análisis.

A lo largo de este documento, se detallarán los procedimientos llevados a cabo para la construcción de la base Jardineria_Staging, incluyendo el diseño de sus tablas, la migración de datos y las validaciones realizadas para garantizar la integridad de la información. Este proceso permite mejorar la manipulación de los datos y facilita futuras consultas y análisis, asegurando que la información esté bien estructurada y libre de inconsistencias.

Para esto cargamos unas pruebas de SQL las cuales evalúan la integridad de los datos, su calidad y que el proceso de ETL haya sido efectivo.

Pruebas Hechas

Prueba de Duplicados:

- Verificación de claves primarias duplicadas en clientes, productos, pedidos y detalle_pedido

```
Pruebas Calidad > pruebas_duplicados.sql
1 SELECT id_cliente, COUNT(*) FROM clientes GROUP BY id_cliente HAVING COUNT(*) > 1;
2 SELECT id_producto, COUNT(*) FROM productos GROUP BY id_producto HAVING COUNT(*) > 1;
3
```

Prueba de Consistencia de Tipos de Datos:

- Revisión de valores no numéricos en campos de tipo DECIMAL, INT.

```
Pruebas Calidad > pruebas_consistencia_datos.sql
1 SELECT * FROM productos WHERE precio NOT LIKE '%[0-9]%';
```

Prueba de Valores Nulos en Campos Clave:

- Verificación de campos clave no nulos.

```
Pruebas Calidad > pruebas_nulos.sql
1 SELECT * FROM pedidos WHERE id_cliente IS NULL;
```

Integridad Referencial:

- Validación de que todas las claves externas tienen correspondencia.

```
Pruebas Calidad > prueba_integridad.sql
1 SELECT dp.* FROM detalle_pedido dp LEFT JOIN pedidos p ON dp.id_pedido = p.id_pedido WHERE p.id_pedido IS NULL;
```

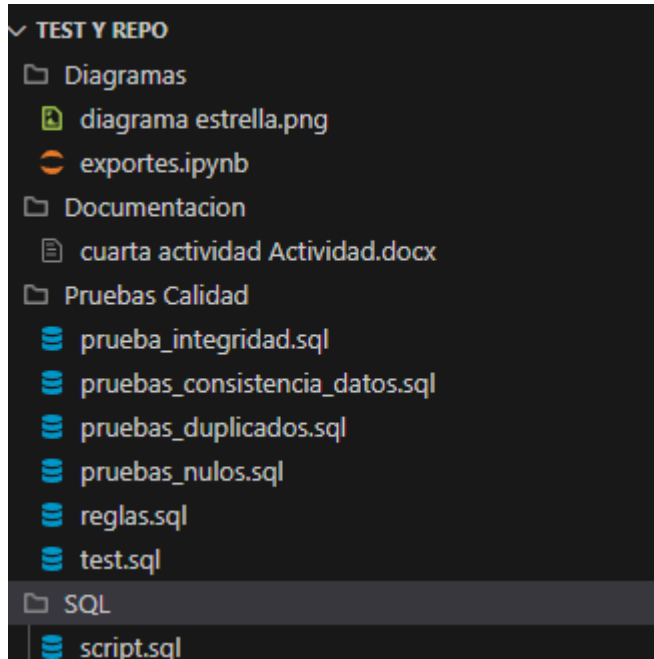
Verificación de Reglas de Negocio:

- Por ejemplo, validar que los subtotales en detalle_pedido coincidan con la multiplicación de cantidad por precio.

```
Pruebas Calidad > reglas.sql
1 SELECT dp.* FROM detalle_pedido dp
2 JOIN producto p ON dp.id_producto = p.id_producto
3 WHERE dp.subtotal <> dp.cantidad * p.precio_venta;
```

2. Organización del Repositorio

2.1 Estructura de Carpetas



Link del repositorio:

<https://github.com/JSebastianCCorrea/BDIICuartaActividad>