

## **Evidencia de aprendizaje #4**

Proyecto integrador-Repositorio de todas las actividades

### **Integrantes Grupo #13**

JOSE LUIS BENITEZ SALCEDO  
GERMAN EDUARDO PAIBA ROCHA  
JOSUÉ SANMARTÍN POLO  
ALEJANDRO CADAVID URREA

**Institución Universitaria Digital de Antioquia**  
**Ingeniería de Software y Datos**

Asignatura:  
**BASE DE DATOS II**

Nombre del tutor:  
**VICTOR HUGO MERCADO**

**Abril 06 de 2025**

## INTRODUCCIÓN

Este documento hace parte de nuestra formación en desarrollo de software, en donde implementamos un proceso ETL completo para transformar los datos operacionales de una base de datos de jardinería en un Data Mart analítico.

Primero diseñamos un modelo estrella con tablas de dimensiones (cliente, producto, tiempo) y hechos (ventas), optimizado para consultas rápidas. Luego se creó una base de datos staging donde se extrajeron, limpiaron y validaron los datos originales, solucionando problemas como valores nulos y formatos inconsistentes.

Finalmente, mediante consultas SQL se transformó y cargó estos datos al Data Mart, generando una estructura lista para análisis empresarial. Este proyecto nos ha permitido aplicar conceptos clave de ingeniería de software y datos - desde el modelado dimensional hasta técnicas ETL - demostrando cómo transformar datos crudos en información valiosa para la toma de decisiones.

## OBJETIVOS

- Unificar todo el trabajo desarrollado en las demás evidencias en una sola entrega.
- Desarrollar un proceso de transformación y carga de datos desde la base de datos origen, pasar Staging y luego hasta el data mart final, utilizando la base de datos de staging previamente creada.
- Asegurar que los datos estén preparados para el análisis, incluyendo la identificación del producto más vendido y otros análisis relevantes para la toma de decisiones empresariales.

## PLANTEAMIENTO DEL PROBLEMA

Creación de repositorio de Git para todas las actividades (incluyendo pruebas de verificación de calidad de datos).

1. Realiza pruebas que te permitan determinar la calidad de los datos.
2. Organiza las actividades:

Reúne todos los archivos y documentos relacionados con cada una de las actividades desarrolladas, incluyendo los informes, consultas SQL, scripts, diagramas y, ahora, también las pruebas de verificación de calidad de datos.

3. Crea una estructura de carpetas

Diseña una estructura de carpetas clara y organizada para almacenar todos los archivos en el repositorio de Git. Asegúrate de incluir una carpeta específica para las pruebas de calidad de datos.

4. Agrega archivos al repositorio
5. Confirma los cambios

## PROPUESTA DE LA SOLUCIÓN

### 1. Creación del Repositorio Git

Se centraliza todo el material del proyecto (scripts, informes, pruebas). Se crea un repositorio nuevo en GitHub y se clona el repositorio localmente:

### 2. Pruebas de Calidad de Datos

Se valida que los datos sean consistentes, completos y correctos.

Pruebas realizadas:

- Integridad: Verificar que no haya valores nulos en campos clave (ej: ID\_CLIENTE, FECHA\_PEDIDO).
- Consistencia: Validar formatos (ej: fechas en YYYY-MM-DD, códigos postales válidos).
- Validación de duplicados: Asegurar que no existan duplicados en tablas críticas.
- Reglas de negocio: Confirmar que los datos cumplen con lógicas específicas (ej: PRECIO\_UNITARIO > 0).

**3. Organización de Archivos:** Se describe la estructura de carpetas propuesta.

**4. Agregar Archivos al Repositorio:** Se mueven todos los archivos a sus carpetas correspondientes.

**5. Verificación Final:** Se valida que todos los scripts SQL funcionan sin errores y los documentos están actualizados, de acuerdo con las entregas realizadas.

Se entrega un repositorio Git organizado que contiene:

Código ejecutable (ETL completo).

Documentación clara (informes, diagramas).

Pruebas de calidad automatizadas o documentadas.

## CONCLUSIONES

La implementación del modelo estrella demostró ser una solución óptima para el análisis de datos, permitiendo consultas rápidas y una estructura clara para las métricas de ventas y dimensiones relacionadas.

El proceso de extracción, transformación y carga (ETL) aseguró datos consistentes y listos para análisis, gracias a la limpieza, validaciones y pruebas de calidad implementadas.

El Data Mart resultante proporciona una base sólida para generar insights empresariales, apoyando la toma de decisiones con datos confiables y accesibles. así mismo, el uso de Git para gestionar scripts, documentación y pruebas facilitó el control de versiones y la colaboración, asegurando un proyecto bien estructurado y replicable.

Con este trabajo, estamos adquiriendo herramientas valiosas que nos ayudarán a destacar en el mercado laboral y contribuir al desarrollo tecnológico del país. Las técnicas aprendidas son escalables a empresas colombianas, las cuales en su mayoría, son ajenas a este tipo de análisis y de uso de la tecnología.

## ANEXOS

Anexo 1. Carpeta comprimida con la información y proyecto.

<https://drive.google.com/drive/folders/1bSpGYKlpGYLXdYIPSCW4p7Km4q8hAtfX?usp=sharing>

Anexo 2: Trabajo Entrega No. 2 Y 3:

[https://drive.google.com/drive/folders/1QKrserYWclbi2tNSJftfm31\\_jfWlghGQ?usp=drive\\_link](https://drive.google.com/drive/folders/1QKrserYWclbi2tNSJftfm31_jfWlghGQ?usp=drive_link)

Anexo 3: Link repositorio Github

[https://github.com/gerpaiba/bases\\_datos\\_II.git](https://github.com/gerpaiba/bases_datos_II.git)

## BIBLIOGRAFÍA

García, F., & Romero, O. (2015). *Data warehousing y business intelligence: Fundamentos, herramientas y aplicaciones*. Editorial UOC.

Hernández Orallo, J. (2004). *Data mining: Conceptos y herramientas*. Pearson Educación.

Inmon, W. H. (2005). *Construcción del data warehouse (3ª ed.)*. Prentice Hall.

Kimball, R., & Ross, M. (2013). *El data warehouse toolkit: Las técnicas definitivas para el diseño dimensional (3ª ed.)*. Wiley. (Traducción al español)

Marqués, J. (2018). *Business intelligence: Análisis y explotación de datos para la toma de decisiones*. Ediciones Paraninfo.

Pérez López, C. (2016). *Business intelligence: Data mining y análisis predictivo*. RC Libros.