

EVIDENCIA DE APRENDIZAJE

CREACIÓN DE UNA BASE DE DATOS DE STAGING

JUAN DAVID CALLE CORREA

EDISON MENA ARROYO

MARIA FERNANDA VILLADA QUINTERO

PROGRAMA

INGENIERÍA DE SOFTWARE Y DATOS

UNIVERSIDAD DIGITAL DE ANTIOQUIA

PREICA2502B010064

CURSO

BASES DE DATOS II

DOCENTE

ANTONIO JESÚS VALDERRAMA

2025-02

Informe de Análisis de Datos – Data Mart Jardinería

Introducción

En este proyecto trabajamos con los datos de la empresa Jardinería, para poder analizarlos de manera más fácil y organizada. La información estaba originalmente en la base de datos operativa, donde se guardan todas las ventas y registros del día a día. Sin embargo, esos datos no estaban listos para hacer análisis directamente, porque están diseñados para registrar transacciones y no para sacar conclusiones rápidas.

Por eso, primero se crearon las tablas en una base de datos intermedia llamada Staging Jardinería, donde se limpió y organizó la información. Una vez que los datos estaban listos, se cargaron en el Data Mart Jardinería, que es una base de datos pensada específicamente para análisis. Allí se crearon las dimensiones (clientes, productos, categorías y tiempo) y la tabla de hechos de ventas, que nos permite ver detalles como cuántos productos se vendieron, a qué clientes y en qué fechas.

Además, se hicieron copias de seguridad de ambas bases de datos para asegurarnos de que la información no se pierda y se pueda restaurar en caso de cualquier problema. Por último, se realizaron consultas para validar que los datos estuvieran correctos y para generar análisis como ventas por cliente, por producto y por mes.

Este proyecto nos permitió ver cómo se transforma información de bases de datos operativas a un Data Mart listo para análisis, y cómo se pueden usar los datos de manera efectiva para tomar decisiones sobre ventas y productos.

Objetivos

Objetivo General

Organizar y analizar los datos de ventas de la empresa Jardinería mediante la construcción de un Data Mart, de manera que se puedan obtener conclusiones claras sobre clientes, productos y tendencias de ventas.

Objetivos Específicos

Revisar y comprender los datos que se encuentran en la base de datos operativa de Jardinería, identificando cuáles son relevantes para el análisis.

Crear una base de datos intermedia (Staging Jardinería) para limpiar, validar y preparar la información antes de cargarla al Data Mart.

Diseñar y construir las dimensiones y la tabla de hechos en el Data Mart para que permitan realizar análisis detallados de ventas.

Cargar los datos desde la base de Staging al Data Mart de manera correcta, asegurando que la información esté completa y sin errores.

Realizar consultas de validación para confirmar que los datos se cargaron correctamente y reflejan la realidad de las ventas.

Generar respaldos (backups) de las bases de datos para garantizar la seguridad y recuperación de la información ante cualquier eventualidad.

Analizar los datos utilizando consultas que muestren ventas por cliente, por producto y por mes, de manera que se puedan obtener conclusiones útiles para la empresa.

Planteamiento del problema

La empresa Jardinería cuenta con muchos datos sobre ventas, clientes y productos, pero la información se encuentra distribuida en la base de datos operativa. Esto genera varias dificultades:

Dificultad para analizar los datos: La base de datos está pensada para registrar transacciones, no para hacer análisis o reportes rápidos.

Falta de información consolidada: Los datos están en diferentes tablas, lo que hace complicado ver, por ejemplo, cuáles productos se venden más o cuáles clientes compran con frecuencia.

Riesgo de errores: Sin una etapa de validación y limpieza, los datos pueden tener errores, duplicados o información incompleta, lo que afectaría los resultados de cualquier análisis.

Toma de decisiones limitada: La empresa necesita tomar decisiones basadas en información confiable y rápida, como identificar tendencias de ventas, planear inventarios o enfocarse en clientes que generan más ingresos.

Por estas razones, es necesario construir un Data Mart que permita:

Organizar los datos de manera estructurada y limpia.

Relacionar las ventas con clientes, productos y tiempo de manera eficiente.

Facilitar el análisis y la generación de reportes para apoyar la toma de decisiones.

Este proyecto busca resolver estos problemas mediante un proceso ordenado: revisar los datos originales, limpiarlos y cargarlos en un Data Mart que facilite el análisis y genere información útil para la empresa.

Análisis del problema

Al observar los datos de la empresa Jardinería, nos dimos cuenta de que no era tan fácil analizarlos como parecía. La información está repartida en varias tablas de la base de datos, y aunque todo está registrado, no está organizada de forma que uno pueda sacar conclusiones rápidamente.

Algunos de los problemas principales que notamos fueron:

Datos desordenados: Las ventas, los productos y los clientes están en diferentes tablas, lo que hace complicado obtener un resumen general de lo que está pasando en la empresa.

Posibles errores o duplicados: Hay información que puede estar incompleta o repetida, y esto podría cambiar los resultados si no se limpia antes de analizar.

Dificultad para hacer análisis rápido: Por ejemplo, no se podía ver fácilmente cuáles productos se venden más, cuáles clientes compran más seguido o cómo varían las ventas mes a mes.

Por todo esto, quedó claro que la mejor solución era construir un Data Mart. Primero usando una base de Staging para organizar y limpiar los datos, y luego creando la base de Data Mart con dimensiones y una tabla de hechos. De esta manera, los datos quedan más fáciles de usar para consultas y análisis, y la empresa puede tomar decisiones basadas en información confiable.

Jardinería (origen) → Staging Jardinería → Data Mart Jardinería

- **Jardinería:** Base de datos operativa con los registros originales de ventas.

- **Staging Jardinería:** Área intermedia para limpieza, validación y transformación de datos.
- **Data Mart Jardinería:** Base de datos orientada a análisis, con tablas de hechos y dimensiones.

Estructura del Data Mart

Tablas de dimensiones:

1. **Dim Cliente** – Información de los clientes.
2. **Dim Producto** – Información de los productos.
3. **Dim Tiempo** – Información temporal (día, mes, año).
4. **Dim Categoría** – Clasificación de productos.

Tabla de hechos:

- **Hechos Ventas** – Registra todas las ventas, incluyendo **Venta ID, Producto ID, Cliente ID, Tiempo ID, Cantidad y Total.**

Consultas de validación y análisis

Ventas totales por cliente

Interpretación: Esta consulta permite identificar los clientes que generan mayores ingresos y validar la correcta relación entre ventas y clientes.

```
SELECT
    c.NombreCliente,
    SUM(h.Total) AS TotalComprado,
    COUNT(h.VentaID) AS NumCompras
FROM HechosVentas h
JOIN DimCliente c ON h.ClienteID = c.ClienteID
GROUP BY c.NombreCliente
ORDER BY TotalComprado DESC;
```

Ventas por producto

Interpretación: Permite determinar qué productos son los más vendidos y si los totales coinciden con las unidades vendidas.

```
SELECT
    p.NombreProducto,
    SUM(h.Total) AS TotalVendido,
    SUM(h.Cantidad) AS TotalUnidades
FROM HechosVentas h
JOIN DimProducto p ON h.ProductoID = p.ProductoID
GROUP BY p.NombreProducto
ORDER BY TotalVendido DESC;
```

Ventas por mes

Interpretación: Muestra la evolución de las ventas a lo largo del tiempo, útil para análisis de tendencias y estacionalidad.

```
SELECT
    CONCAT(t.Anio, '-', t.Mes) AS Mes,
    SUM(h.Total) AS TotalMensual
FROM HechosVentas h
JOIN DimTiempo t ON h.TiempoID = t.TiempoID
GROUP BY t.Anio, t.Mes
ORDER BY t.Anio, t.Mes;
```

productos por cliente

Interpretación: Identifica los productos favoritos de cada cliente y valida la consistencia de las relaciones entre clientes y productos.

```
WITH Ranking AS (
    SELECT
        c.NombreCliente,
        p.NombreProducto,
        SUM(h.Total) AS TotalGastado,
        ROW_NUMBER() OVER (PARTITION BY c.ClienteID ORDER BY SUM(h.Total) DESC) AS Rn
    FROM HechosVentas h
    JOIN DimCliente c ON h.ClienteID = c.ClienteID
    JOIN DimProducto p ON h.ProductoID = p.ProductoID
    GROUP BY c.ClienteID, c.NombreCliente, p.ProductoID, p.NombreProducto
)
SELECT *
FROM Ranking
WHERE Rn <= 5
ORDER BY NombreCliente, TotalGastado DESC;
```

Validación de datos

Para comprobar la calidad de la migración, se realizaron consultas de verificación en ambas bases. Estas pruebas consistieron en contar el número de registros por tabla y en calcular sumas de valores clave, con el fin de contrastar los resultados entre la base Jardinería (origen) y la base Staging Jardineria (destino).

Hallazgos principales:

La cantidad de registros coincide en todas las tablas transferidas: clientes, productos, pedidos, categorías, pagos y detalle de pedidos.

Los indicadores de negocio también mantienen coherencia:

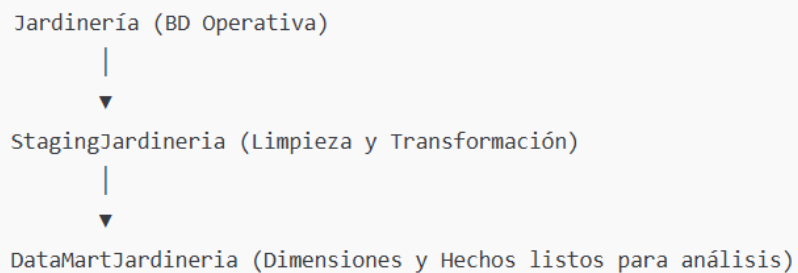
Total, de ventas: 217,738.00

Total, de pagos: 197,940.00

Conclusión:

Con base en estas validaciones, se confirma que la información fue migrada de manera completa y sin inconsistencias hacia la base de datos Staging, garantizando la integridad de los datos y la correcta ejecución del proceso.

Flujo de Datos



Propuesta de la solución

Para mejorar la entrega inicial del proyecto y asegurar que los datos estuvieran correctamente organizados y listos para análisis, se implementaron varias acciones:

Correcciones a la entrega 1

- Se revisaron los nombres de tablas y columnas para corregir errores de escritura y garantizar consistencia.
- Se ajustaron las relaciones entre las tablas de dimensiones y la tabla de hechos, evitando problemas de integridad referencial.
- Se verificó que los datos de ventas, clientes y productos coincidieran con los registros originales de la base de datos de Jardinería.

Traslado a Data Mart y análisis

Una vez validados en Staging, los datos se cargaron al Data Mart Jardinería, construyendo la tabla de hechos Hechos Ventas y las dimensiones Dim Cliente, Dim Producto, Dim Categoría y Dim Tiempo.

Se realizaron consultas de validación para asegurar que los totales de ventas y las cantidades coincidieran con los datos originales.

Se pudieron generar análisis de ventas por cliente, por producto y por mes, permitiendo obtener información útil para la empresa.

Crear la base de datos y usarla

```
-- Crear la base de datos Staging
CREATE DATABASE StagingJardineria;
GO

-- Usar la base de datos recién creada
USE StagingJardineria;
GO
```

Crear la tabla de clientes

```
-- Tabla de clientes
CREATE TABLE Stg_Clientes (
    ClienteID INT PRIMARY KEY,
    NombreCliente NVARCHAR(100),
    Ciudad NVARCHAR(50),
    Telefono NVARCHAR(20)
);
```

Crear la tabla de categorías

```
-- Tabla de categorías
CREATE TABLE Stg_Categorias (
    CategoriaID INT PRIMARY KEY,
    NombreCategoria NVARCHAR(50)
);
```

Crear la tabla de productos

```
-- Tabla de productos
CREATE TABLE Stg_Productos (
    ProductoID INT PRIMARY KEY,
    NombreProducto NVARCHAR(100),
    CategoriaID INT,
    PrecioUnitario DECIMAL(10,2),
    FOREIGN KEY (CategoriaID) REFERENCES Stg_Categorias(CategoriaID)
);
```

Crear la tabla de ventas

```
-- Tabla de ventas
CREATE TABLE Stg_Ventas (
    VentaID INT PRIMARY KEY,
    ClienteID INT,
    ProductoID INT,
    TiempoID INT,
    Cantidad INT,
    Total DECIMAL(10,2),
    FOREIGN KEY (ClienteID) REFERENCES Stg_Clientes(ClienteID),
    FOREIGN KEY (ProductoID) REFERENCES Stg_Productos(ProductoID)
);
```

Carga de datos desde Jardinería hacia Staging

```
1. Categorías
===== */
INSERT INTO dbo.StgCategoria (CategoriaID, NombreCategoria)
SELECT Id_Categoria, Desc_Categoria
FROM Jardineria.dbo.Categoria_producto;

/* =====
2. Clientes
===== */
INSERT INTO dbo.StgCliente (ClienteID, Nombre, Ciudad, Pais)
SELECT ID_cliente, nombre_cliente, ciudad, pais
FROM Jardineria.dbo.Cliente;

/* =====
3. Productos
===== */
INSERT INTO dbo.StgProducto (ProductoID, Nombre, CategoriaID, Precio)
SELECT ID_producto, nombre, Categoria, precio_venta
FROM Jardineria.dbo.Producto;
```

```

4. Pedidos
===== */
INSERT INTO dbo.StgPedido (PedidoID, ClienteID, FechaPedido, FechaEsperada, FechaEntrega, Est
SELECT ID_pedido, ID_cliente, fecha_pedido, fecha_esperada, fecha_entrega, estado
FROM Jardineria.dbo.Pedido;

/* =====
5. Detalle de Pedido
===== */
INSERT INTO dbo.StgDetallePedido (PedidoID, ProductoID, Cantidad, PrecioUnidad)
SELECT ID_pedido, ID_producto, cantidad, precio_unidad
FROM Jardineria.dbo.Detalle_pedido;

/* =====
6. Pagos
===== */
INSERT INTO dbo.StgPago (PagoID, ClienteID, FechaPago, Total)
SELECT ID_pago, ID_cliente, fecha_pago, total
FROM Jardineria.dbo.Pago;

```

Validación por conteo de registros

```

-- Categorías
SELECT COUNT(*) AS Total_Origen FROM Jardineria.dbo.Categoria_producto;
SELECT COUNT(*) AS Total_Staging FROM StagingJardineria.dbo.StgCategoria;

-- Clientes
SELECT COUNT(*) AS Total_Origen FROM Jardineria.dbo.Cliente;
SELECT COUNT(*) AS Total_Staging FROM StagingJardineria.dbo.StgCliente;

-- Productos
SELECT COUNT(*) AS Total_Origen FROM Jardineria.dbo.Producto;
SELECT COUNT(*) AS Total_Staging FROM StagingJardineria.dbo.StgProducto;

```

```
-- Pedidos
SELECT COUNT(*) AS Total_Origen FROM Jardineria.dbo.Pedido;
SELECT COUNT(*) AS Total_Staging FROM StagingJardineria.dbo.StgPedido;

-- Detalle de Pedido
SELECT COUNT(*) AS Total_Origen FROM Jardineria.dbo.Detalle_pedido;
SELECT COUNT(*) AS Total_Staging FROM StagingJardineria.dbo.StgDetallePedido;

-- Pagos
SELECT COUNT(*) AS Total_Origen FROM Jardineria.dbo.Pago;
SELECT COUNT(*) AS Total_Staging FROM StagingJardineria.dbo.StgPago;
```

Validación de negocio

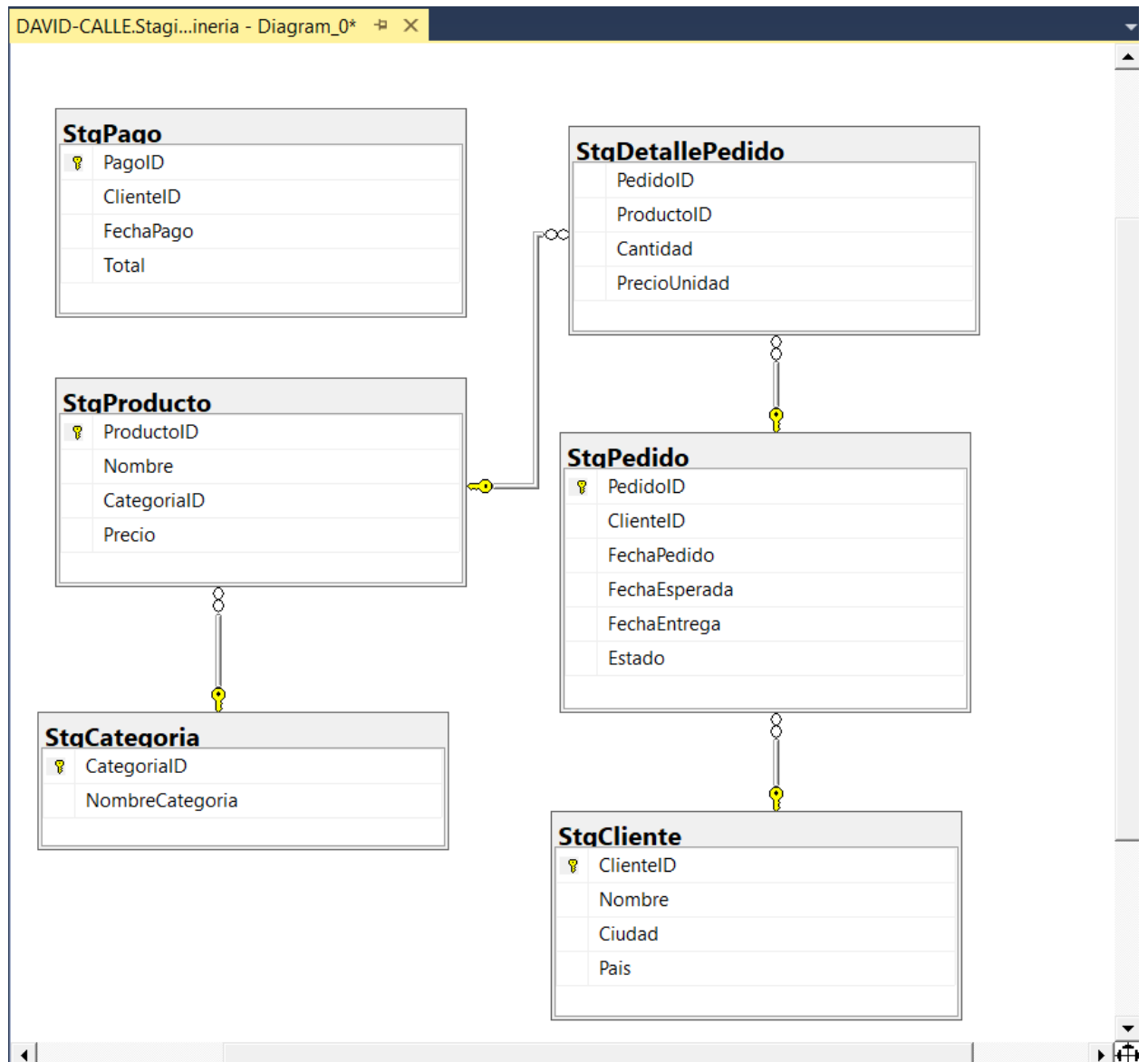
```
-- Total de ventas en Jardineria (origen)
SELECT SUM(cantidad * precio_unidad) AS Total_Ventas_Origen
FROM Jardineria.dbo.Detalle_pedido;

-- Total de ventas en Staging
SELECT SUM(Cantidad * PrecioUnidad) AS Total_Ventas_Staging
FROM StagingJardineria.dbo.StgDetallePedido;

-- Total de pagos en Jardineria (origen)
SELECT SUM(total) AS Total_Pagos_Origen
FROM Jardineria.dbo.Pago;

-- Total de pagos en Staging
SELECT SUM(Total) AS Total_Pagos_Staging
FROM StagingJardineria.dbo.StgPago;
```

Modelo Staging



Backups realizados

Para garantizar la seguridad y recuperación de los datos, se realizaron copias de respaldo de las bases de datos involucradas en el proyecto:

Staging Jardinería

```
-- Backup completo de la base de datos Jardineria
BACKUP DATABASE Jardineria
TO DISK = 'C:\Backups\Jardineria.bak'
WITH INIT,
    NAME = 'Backup completo Jardineria',
    STATS = 10;
GO
```

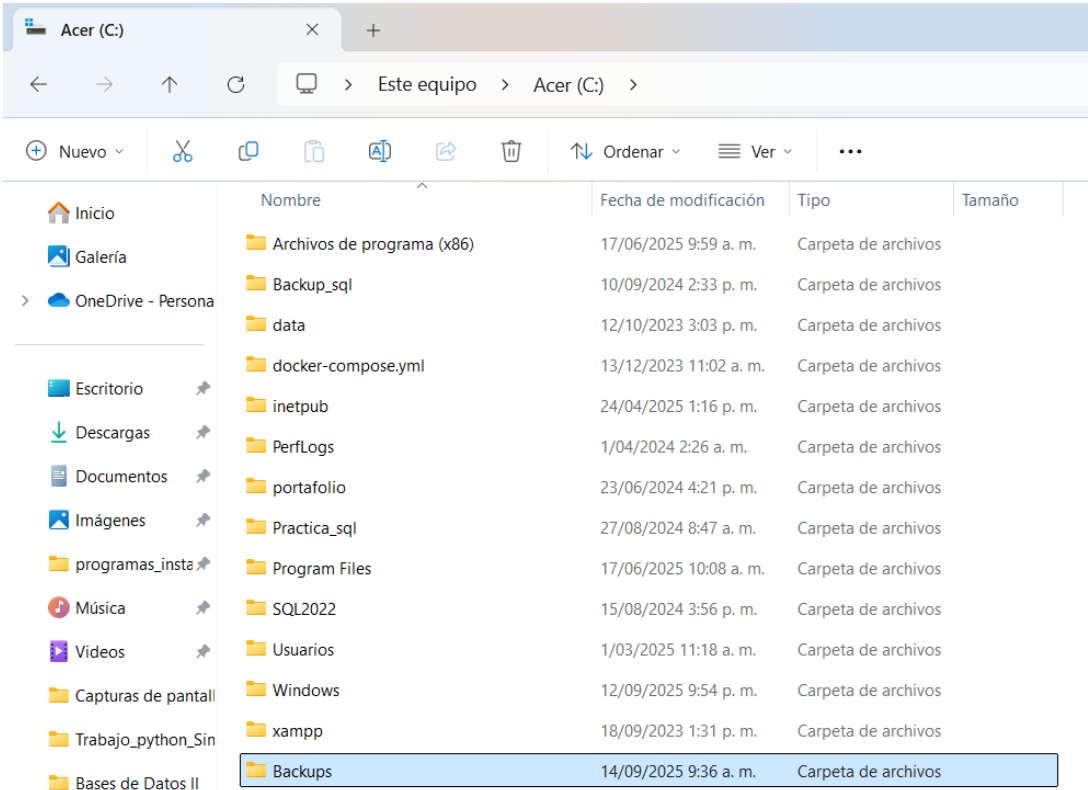
- **Propósito:** Proteger los datos temporales transformados antes de cargarlos al Data Mart.
- **Resultado:** Backup completado exitosamente, permitiendo restauración ante cualquier fallo.

Backup de StagingJardineria

```
-- Backup completo de la base de datos StagingJardineria
BACKUP DATABASE StagingJardineria
TO DISK = 'C:\Backups\StagingJardineria.bak'
WITH INIT,
    NAME = 'Backup completo StagingJardineria',
    STATS = 10;
GO
```

- **Propósito:** Asegurar la integridad de los datos finales de análisis y reportes.
- **Resultado:** Backup completado correctamente, garantizando que toda la información consolidada esté protegida.

Nota: Los backups se recomiendan ejecutar regularmente y almacenarse en un lugar seguro, idealmente fuera del servidor de producción, para recuperación rápida en caso de incidentes.



Bibliografía

Inmon, W. H. (2005). Construyendo el data warehouse corporativo. México: McGraw-Hill.

Esta fuente explica los conceptos de Data Warehousing, la construcción de tablas de hechos y dimensiones, y el flujo de datos desde sistemas operativos a almacenes de datos.

Kimball, R., & Ross, M. (2013). El Data Warehouse Toolkit: Guía para diseño dimensional (3.^a ed.). España: Wiley.

Referencia utilizada para la creación de tablas de hechos, dimensiones y la metodología ETL (Extracción, Transformación y Carga).

Hernández, R., Fernández, C., & Baptista, P. (2014). Metodología de la investigación (6.^a ed.). México: McGraw-Hill.

Se consultó para el planteamiento del problema, análisis de datos y redacción de objetivos del proyecto.

Date, C. J. (2010). Introducción a los sistemas de bases de datos (8.^a ed.). España: Pearson Educación.

Apoyo en la definición de estructuras de bases de datos, claves primarias y foráneas, y relaciones entre tablas.

Gutiérrez, E. (2018). ETL y transformación de datos en entornos empresariales. Bogotá: Editorial Universidad Nacional.

Fuente utilizada para explicar el proceso de traslado de datos desde la base operativa a Staging y luego al Data Mart.

SQL Server Docs. (s.f.). Backup y restauración de bases de datos. Microsoft Docs.

Recuperado de <https://learn.microsoft.com/es-es/sql/relational-databases/backup-restore/back-up-and-restore-of-sql-server-databases>

Consultado para la creación de backups de las bases de datos Staging y Data Mart.