

ACTIVIDAD DE APRENDIZAJE 4.

Carlos Arturo Gómez Jiménez

Grupo 16

Ingeniería de Software y Datos

Facultad de Ingeniería y Ciencias Agropecuarias

Institución Universitaria Digital de Antioquia

Bases de Datos II

Ing. Victor Hugo Mercado

06 de abril del 2025



TABLA DE CONTENIDO

1. INTRODUCCIÓN	3
2. OBJETIVOS.....	3
OBJETIVOS ESPECÍFICOS	4
3. PLANTEAMIENTO DEL PROBLEMA	4
4. ANÁLISIS DEL PROBLEMA	5
5. PROPUESTA DE LA SOLUCIÓN	6
DESCRIPCIÓN DEL MODELO ESTRELLA PROPUESTO	6
RELACIÓN ENTRE LA TABLA DE HECHOS Y LAS DIMENSIONES	7
DISEÑO DEL MODELO ESTRELLA	7
6. DISEÑO DE LA BASE DE DATOS STAGING	8
DESCRIPCIÓN DEL MODELO ESTRELLA PROPUESTO	8
OBJETIVO DEL ENTORNO DE STAGING.....	9
ESTRUCTURA DE STAGING	9
PROCESO DE EXTRACCIÓN DE DATOS.....	9
TRANSFORMACIÓN DE DATOS	9
CARGA FINAL AL DATA MART	9
VALIDACIÓN DE DATOS.....	10
7. CONSULTAS SQL IMPLEMENTADAS	10
CREACIÓN DE TABLAS EN STAGING	10
EXTRACCIÓN DE DATOS	10
TRANSFORMACIONES.....	10
CARGA AL DATA MART	11
VALIDACIONES.....	11
8. CONCLUSIONES	11
9. BIBLIOGRAFÍA.....	12

1. INTRODUCCIÓN

En la actualidad, la gestión eficiente de datos es fundamental para el desarrollo estratégico y operativo de las organizaciones. La capacidad de analizar información histórica y transaccional de manera estructurada permite detectar patrones, optimizar procesos y tomar decisiones informadas.

Dentro de este contexto, el presente proyecto propone la implementación de un modelo de datos en forma de estrella aplicado a la base de datos de una empresa dedicada al comercio de productos de jardinería. Esta empresa gestiona información relacionada con pedidos, clientes, productos, empleados y oficinas de atención, siendo crucial centralizar y estructurar estos datos para su análisis posterior.

El modelo estrella, ampliamente utilizado en sistemas de inteligencia de negocios, se caracteriza por una tabla de hechos central que registra las transacciones o eventos clave del negocio (en este caso, los pedidos), rodeada por varias tablas de dimensiones que describen los distintos elementos involucrados en esas transacciones (clientes, productos, tiempo, entre otros). Esta estructura facilita la consulta eficiente de grandes volúmenes de datos, optimiza los tiempos de respuesta y permite la elaboración de reportes analíticos claros y segmentados.

A través del diseño de este modelo, se busca transformar un sistema de datos transaccional en una estructura orientada al análisis, apoyada por un proceso de extracción, transformación y carga (ETL). Este proceso incluye la creación de un entorno de staging intermedio, la depuración y normalización de los datos, y la posterior carga en un Data Mart listo para ser explotado por herramientas de inteligencia empresarial (BI).

2. OBJETIVOS

El objetivo principal de este proyecto es implementar un modelo de datos tipo estrella que facilite el análisis eficiente de la información de ventas de una empresa de jardinería. A través de este modelo, se busca organizar los datos transaccionales en una estructura optimizada para consultas analíticas, mejorando la disponibilidad, calidad y usabilidad de la información.

OBJETIVOS ESPECÍFICOS

- Diseñar una tabla de hechos central que consolide las transacciones de ventas de la empresa, capturando las métricas clave de cada pedido, como la cantidad vendida, el precio por unidad, el total de la transacción y las fechas asociadas.
- Definir y construir las dimensiones relevantes que describen los principales actores del negocio, incluyendo clientes, productos, empleados, oficinas y tiempo, permitiendo una segmentación detallada de la información para su análisis.
- Establecer relaciones claras entre la tabla de hechos y las dimensiones, garantizando la integridad referencial y la coherencia lógica del modelo, con el fin de permitir una navegación fluida en el análisis de datos multidimensional.
- Implementar un proceso de ETL estructurado, que permita la extracción de datos desde la base transaccional original, su transformación (limpieza, normalización y enriquecimiento), y la posterior carga en un Data Mart orientado a la explotación analítica.
- Optimizar el rendimiento del sistema de consultas mediante el diseño de un modelo que minimice la complejidad de las consultas SQL, mejore los tiempos de respuesta y facilite la generación de reportes para la toma de decisiones estratégicas.

3. PLANTEAMIENTO DEL PROBLEMA

En la empresa de jardinería, los datos relacionados con ventas, clientes, productos y empleados se encuentran almacenados en un sistema transaccional cuya estructura no está optimizada para el análisis ni la toma de decisiones estratégicas. Las consultas requeridas para obtener información clave sobre el comportamiento de las ventas o la relación entre los distintos actores del negocio implican el cruce de múltiples tablas con estructuras complejas, lo que genera una elevada carga computacional y tiempos de respuesta prolongados.

Este entorno dificulta la elaboración de reportes ágiles y precisos, ya que carece de una arquitectura orientada al análisis multidimensional. La ausencia de un modelo centralizado que conecte

adecuadamente los datos impide identificar tendencias, segmentar clientes por comportamiento o realizar evaluaciones detalladas del desempeño de los productos y del personal de ventas.

Además, la falta de un sistema de integración y limpieza de datos genera inconsistencias que afectan la calidad de la información utilizada en el análisis. En consecuencia, los tomadores de decisiones no cuentan con una fuente confiable ni con una estructura que facilite el acceso a los indicadores más relevantes del negocio.

4. ANÁLISIS DEL PROBLEMA

El análisis del problema revela que la estructura actual de la base de datos responde a un modelo puramente transaccional, centrado en registrar eventos operativos, pero sin considerar su uso posterior en procesos analíticos. Esta arquitectura, aunque adecuada para la operación diaria, presenta limitaciones significativas al momento de ejecutar consultas complejas que involucren múltiples entidades, como clientes, productos, empleados, oficinas y fechas.

Entre las principales deficiencias detectadas se encuentran:

- Fragmentación de la información: los datos relevantes para el análisis de ventas están distribuidos en varias tablas sin una estructura analítica definida, lo que dificulta la integración y el acceso coherente a la información.
- Consultas poco eficientes: las uniones necesarias entre tablas para obtener reportes relevantes afectan negativamente el rendimiento del sistema, lo cual impacta la productividad de los usuarios encargados del análisis.
- Falta de integración semántica: las relaciones entre las distintas entidades del negocio no están modeladas de forma que faciliten la interpretación y explotación de los datos desde una perspectiva de inteligencia de negocios.
- Ausencia de un esquema de tiempo estructurado: el análisis por periodos, trimestres, años o comparativos temporales resulta complicado debido a la carencia de una dimensión de tiempo formalmente definida.

- Riesgos de inconsistencia: la falta de un proceso sistemático de transformación y validación previo al análisis puede generar errores o decisiones basadas en datos incorrectos o incompletos.

5. PROPUESTA DE LA SOLUCIÓN

Para resolver las limitaciones identificadas en el análisis del problema, se propone el diseño e implementación de un modelo de datos tipo estrella, una arquitectura ampliamente utilizada en almacenes de datos por su simplicidad, eficiencia y orientación al análisis multidimensional.

El modelo estrella se basa en una tabla de hechos central que almacena las transacciones clave del negocio (en este caso, los pedidos realizados por los clientes), y un conjunto de tablas de dimensiones que describen los distintos elementos involucrados en dichas transacciones. Esta estructura permite realizar consultas rápidas, claras y segmentadas sobre grandes volúmenes de datos.

DESCRIPCIÓN DEL MODELO ESTRELLA PROPUESTO

Tabla de Hechos: fact_pedido

Esta tabla representa las transacciones de ventas realizadas por la empresa. Cada registro corresponde a una línea de pedido e incluye información cuantitativa como la cantidad de productos vendidos, el precio unitario y el total de la venta. Además, contiene claves foráneas que enlazan con las dimensiones descriptivas:

- ID_cliente
- ID_empleado
- ID_producto
- ID_oficina
- fecha_pedido

Tablas de Dimensiones

- dim_cliente: almacena información sobre los clientes, incluyendo nombre, datos de contacto y localización geográfica.

- **dim_empleado:** contiene los datos del personal de ventas o atención, tales como nombres, apellidos, puesto y oficina asignada.
- **dim_producto:** describe los productos vendidos, sus códigos, nombres, categoría y precio de venta.
- **dim_oficina:** representa las oficinas comerciales de la empresa, con datos de ubicación como ciudad y país.
- **dim_categoria:** clasifica los productos según su categoría general, permitiendo segmentaciones adicionales.
- **dim_tiempo:** estructura las fechas asociadas a los pedidos en unidades de análisis como día, mes, trimestre y año.

RELACIÓN ENTRE LA TABLA DE HECHOS Y LAS DIMENSIONES

La tabla **fact_pedido** establece relaciones mediante claves foráneas con cada dimensión. Esta relación facilita la ejecución de consultas analíticas como:

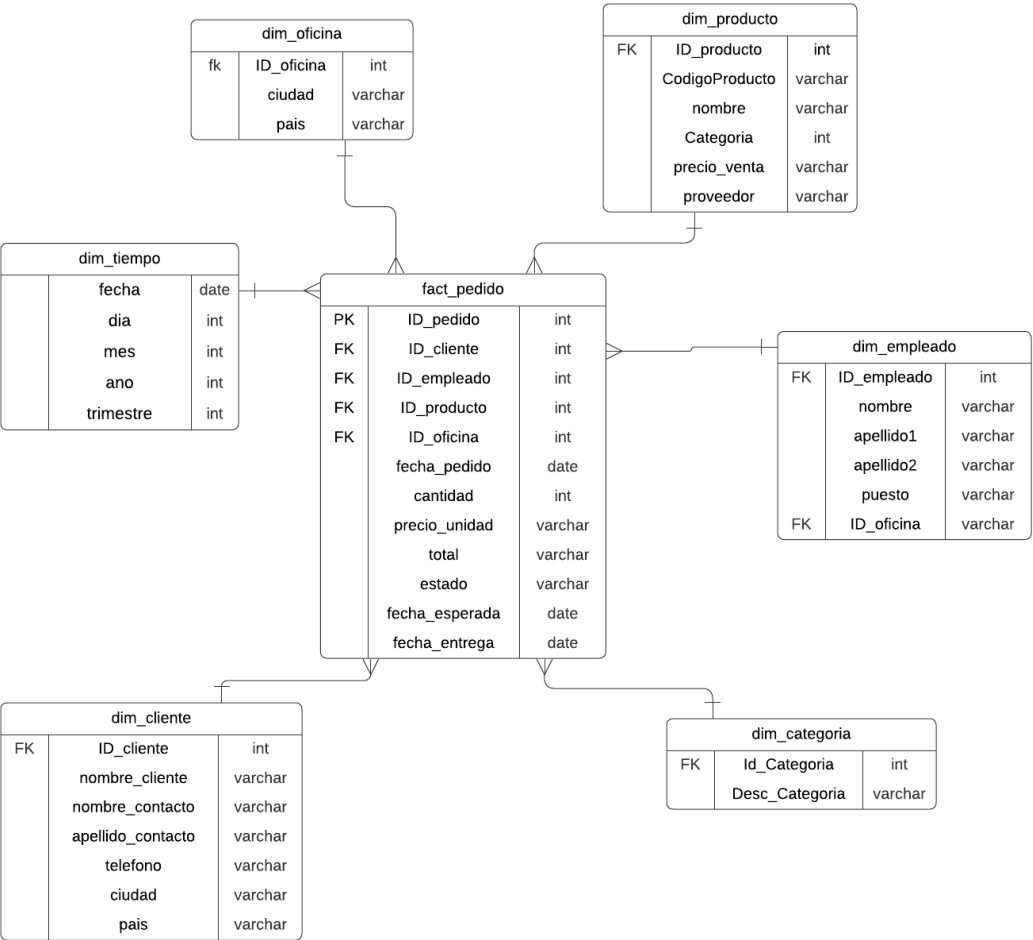
- Ventas por cliente o por país.
- Análisis del rendimiento por empleado o por oficina.
- Comparativas de ventas por producto o categoría.
- Tendencias y patrones de comportamiento por periodos de tiempo.

Relaciones clave:

- ID_cliente → dim_cliente
- ID_empleado → dim_empleado
- ID_producto → dim_producto
- ID_oficina → dim_oficina
- fecha_pedido → dim_tiempo

DISEÑO DEL MODELO ESTRELLA

A continuación, se presenta el diseño gráfico del modelo estrella propuesto, que resume visualmente la estructura lógica entre la tabla de hechos y las dimensiones:



6. DISEÑO DE LA BASE DE DATOS STAGING

DESCRIPCIÓN DEL MODELO ESTRELLA PROPUESTO

La implementación del modelo estrella requiere un paso intermedio para preparar los datos antes de ser cargados en el Data Mart final. Esta etapa es conocida como staging, y cumple un rol clave en el proceso ETL (Extracción, Transformación y Carga), ya que permite controlar la calidad de los datos y aplicar transformaciones sin afectar directamente las fuentes de origen ni el entorno de análisis.

OBJETIVO DEL ENTORNO DE STAGING

El propósito principal de la base de datos staging es actuar como una zona temporal de preparación, donde se almacenan los datos extraídos desde el sistema transaccional original. En este entorno, los datos pueden ser limpiados, transformados, validados y enriquecidos antes de ser trasladados a las tablas analíticas definitivas.

ESTRUCTURA DE STAGING

La base de datos staging replica la estructura lógica del modelo estrella, pero en un entorno temporal. Se diseñaron las siguientes tablas:

- Tabla de hechos: Staging.fact_pedido
- Tablas de dimensiones: Staging.dim_cliente, Staging.dim_empleado, Staging.dim_producto, Staging.dim_oficina, Staging.dim_categoria, Staging.dim_tiempo

PROCESO DE EXTRACCIÓN DE DATOS

Los datos se extrajeron de la base de datos Jardineria mediante consultas INSERT INTO ... SELECT. Este proceso trasladó la información cruda a las tablas de staging sin aplicar transformaciones significativas en esta fase.

TRANSFORMACIÓN DE DATOS

En la etapa de staging, se aplicaron transformaciones para mejorar la calidad y consistencia de los datos. Entre las más relevantes:

- Eliminación de espacios en blanco.
- Conversión de cadenas a mayúsculas.
- Redondeo de valores monetarios.
- Generación de la dimensión dim_tiempo a partir de fechas de pedidos.

CARGA FINAL AL DATA MART

Una vez transformados y validados, los datos fueron cargados en el esquema analítico dm_jardineria. Se utilizó la misma estructura que en staging, pero con los datos ya consolidados y listos para el análisis.

VALIDACIÓN DE DATOS

Después de la carga, se realizaron consultas de validación para asegurar:

- La cantidad de registros esperada en cada tabla.
- La integridad referencial entre hechos y dimensiones.
- La coherencia de cálculos numéricos, como los totales en ventas.

7. CONSULTAS SQL IMPLEMENTADAS

A continuación, se presentan las principales consultas SQL desarrolladas a lo largo del proyecto, agrupadas en función de la fase del proceso ETL en la que fueron aplicadas. Estas sentencias reflejan la lógica del flujo de datos desde la base transaccional hasta el modelo estrella implementado en el Data Mart.

CREACIÓN DE TABLAS EN STAGING

En esta fase se implementan las estructuras temporales necesarias para recibir los datos extraídos desde el sistema transaccional. Las tablas de staging replican la lógica del modelo estrella e incluyen definiciones de claves primarias y relaciones foráneas que garantizan la consistencia de los datos durante el proceso de preparación. Estas estructuras permiten trabajar los datos sin afectar directamente la base origen ni el Data Mart final.

EXTRACCIÓN DE DATOS

Se ejecutan las sentencias que permiten trasladar los datos desde las tablas transaccionales originales hacia las tablas staging. Esta extracción contempla tanto datos dimensionales como transaccionales, y en el caso de la tabla de hechos, se integran varias fuentes mediante uniones lógicas para consolidar la información necesaria. El objetivo de esta etapa es disponer de los datos en bruto, listos para ser depurados y transformados.

TRANSFORMACIONES

En esta etapa se aplican procesos de limpieza, estandarización y normalización de datos sobre las tablas staging. Esto incluye la eliminación de caracteres innecesarios, la conversión de textos a formatos consistentes, el redondeo de valores y la generación de estructuras derivadas, como la dimensión de

tiempo. Estas transformaciones aseguran la calidad, integridad y coherencia de los datos antes de ser cargados en el entorno analítico.

CARGA AL DATA MART

Finalizadas las transformaciones, los datos se trasladan desde staging hacia el esquema definitivo del Data Mart. Esta carga se realiza tabla por tabla, replicando la estructura del modelo estrella. Las dimensiones se cargan primero para garantizar la disponibilidad de referencias, y luego se inserta la tabla de hechos, estableciendo así las relaciones necesarias para el análisis multidimensional.

VALIDACIONES

Para asegurar la correcta ejecución del proceso, se ejecutan consultas de verificación que comprueban la cantidad de registros cargados, la integridad referencial entre hechos y dimensiones, y la ausencia de datos incompletos o inconsistentes. Estas validaciones permiten confirmar que los datos están listos para ser utilizados en consultas analíticas y visualizaciones.

8. CONCLUSIONES

La implementación del modelo estrella propuesto permitió transformar una estructura de base de datos transaccional en un entorno analítico optimizado para la toma de decisiones. A través del proceso ETL desarrollado, se logró consolidar la información dispersa en distintas entidades, estructurarla en torno a una tabla de hechos central y establecer dimensiones bien definidas que permiten un análisis segmentado, eficiente y escalable.

El uso de una base staging como zona intermedia resultó fundamental para garantizar la calidad de los datos mediante procesos de validación, limpieza y transformación. Esto permitió asegurar que los datos cargados en el Data Mart fueran consistentes, completos y preparados para consultas analíticas avanzadas.

El modelo desarrollado mejora sustancialmente el rendimiento en la generación de reportes, facilita el acceso a métricas clave del negocio y habilita la posibilidad de integrar herramientas de inteligencia de negocios. La estructura obtenida permite realizar análisis por cliente, producto, empleado, ubicación y periodo de tiempo, fortaleciendo la capacidad analítica de la organización.

9. BIBLIOGRAFÍA

Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling (3rd ed.). John Wiley & Sons.

Connolly, T., & Begg, C. (2014). Database Systems: A Practical Approach to Design, Implementation, and Management (6th ed.). Pearson Education.

Inmon, W. H. (2005). Building the Data Warehouse (4th ed.). John Wiley & Sons.