

Lección 4: Modelamiento Multidimensional

1. Selección del Área de Negocio

Para el diseño del modelo multidimensional, nos enfocaremos en el **área de Ventas**, que representa un proceso crítico para la empresa de e-commerce y retail digital. Esta área integra datos de múltiples fuentes que ya hemos identificado en las etapas anteriores de la arquitectura, y genera información esencial para la toma de decisiones estratégicas.

1.1. Justificación de la Selección

- **Impacto directo en ingresos:** Las ventas representan la principal fuente de ingresos y un indicador clave del rendimiento del negocio.
- **Necesidades analíticas frecuentes:** Los análisis de ventas son requeridos constantemente por múltiples departamentos (Comercial, Marketing, Finanzas).
- **Riqueza de dimensiones:** Permite analizar el comportamiento desde múltiples perspectivas (clientes, productos, tiempo, ubicación, canales).
- **Disponibilidad de datos:** Las fuentes para alimentar este modelo ya están disponibles en las zonas Trusted y Curated del Data Lake.

1.2. Preguntas de Negocio a Responder

El modelo dimensional debe permitir responder preguntas clave como:

1. ¿Cuáles son las ventas por categoría de producto en cada región y período?
2. ¿Cómo evoluciona la conversión de visitas a compras por canal y segmento de cliente?
3. ¿Qué productos tienen mayor frecuencia de compra repetida?
4. ¿Cuál es el impacto de las promociones en las ventas por categoría?
5. ¿Cuáles son los patrones estacionales de venta por tipo de producto?
6. ¿Cómo se compara el rendimiento entre tiendas físicas y canales digitales?

2. Diseño del Modelo Dimensional

Para el área de Ventas, implementaremos un **esquema estrella** que permitirá análisis eficientes mientras mantiene una estructura clara y comprensible.

2.1. Tabla de Hechos Principal: FACT_VENTAS

La tabla de hechos central capturará las transacciones de venta realizadas a través de todos los canales (online y tiendas físicas).

Columnas:

- **venta_id** (PK): Identificador único de la transacción
- **fecha_id** (FK): Referencia a dimensión tiempo
- **producto_id** (FK): Referencia a dimensión producto
- **cliente_id** (FK): Referencia a dimensión cliente
- **tienda_id** (FK): Referencia a dimensión tienda (física o virtual)
- **promocion_id** (FK): Referencia a dimensión promoción
- **empleado_id** (FK): Referencia a dimensión empleado (para ventas físicas)
- **metodo_pago_id** (FK): Referencia a dimensión método de pago

Métricas (Medidas):

- **cantidad:** Número de unidades vendidas
- **precio_unitario:** Precio por unidad
- **monto_bruto:** Monto total antes de descuentos
- **monto_descuento:** Valor del descuento aplicado

- **monto_impuesto:** Valor del impuesto
- **monto_neto:** Monto final después de descuentos e impuestos
- **costo_unitario:** Costo por unidad (para cálculo de márgenes)
- **margen:** Diferencia entre precio y costo

2.2. Tabla de Hechos Secundaria: FACT_ACTIVIDAD_WEB

Para analizar el comportamiento previo a la compra y las conversiones, incluiremos una tabla de hechos complementaria que captura la actividad web.

Columnas:

- **actividad_id** (PK): Identificador único de la actividad
- **fecha_id** (FK): Referencia a dimensión tiempo
- **producto_id** (FK): Referencia a dimensión producto (si aplica)
- **cliente_id** (FK): Referencia a dimensión cliente (si está identificado)
- **sesion_id:** Identificador de la sesión web
- **pagina_id** (FK): Referencia a dimensión página web
- **dispositivo_id** (FK): Referencia a dimensión dispositivo

Métricas (Medidas):

- **tiempo_pagina:** Tiempo en segundos en la página
- **clicks:** Número de clicks realizados
- **flag_carrito_add:** Indicador de adición al carrito (0/1)
- **flag_carrito_remove:** Indicador de eliminación del carrito (0/1)
- **flag_wishlist:** Indicador de adición a lista de deseos (0/1)
- **flag_checkout_inicio:** Indicador de inicio de checkout (0/1)
- **flag_checkout_completo:** Indicador de checkout completado (0/1)

2.3. Dimensiones Principales

DIM_TIEMPO

- **fecha_id** (PK): Identificador único de fecha
- **fecha:** Fecha completa (YYYY-MM-DD)
- **dia:** Día del mes (1-31)
- **dia_semana:** Día de la semana (Lunes-Domingo)
- **semana:** Número de semana del año (1-53)
- **mes:** Número de mes (1-12)
- **nombre_mes:** Nombre del mes (Enero-Diciembre)
- **trimestre:** Trimestre (1-4)
- **semestre:** Semestre (1-2)
- **año:** Año (YYYY)
- **es_feriado:** Indicador de día feriado (Sí/No)
- **es_fin_semana:** Indicador de fin de semana (Sí/No)
- **temporada:** Temporada comercial (Regular, Navidad, Black Friday, etc.)

DIM_PRODUCTO

- **producto_id** (PK): Identificador único de producto
- **sku:** Código SKU del producto
- **nombre_producto:** Nombre completo del producto
- **descripcion:** Descripción del producto
- **categoria_id:** ID de categoría
- **nombre_categoria:** Nombre de la categoría
- **subcategoria_id:** ID de subcategoría
- **nombre_subcategoria:** Nombre de la subcategoría

- **marca_id**: ID de la marca
- **nombre_marca**: Nombre de la marca
- **proveedor_id**: ID del proveedor
- **nombre_proveedor**: Nombre del proveedor
- **precio_lista**: Precio de lista actual
- **costo**: Costo actual
- **peso**: Peso del producto
- **dimensiones**: Dimensiones del producto
- **color**: Color principal
- **fecha_introduccion**: Fecha de introducción al catálogo
- **estado**: Estado del producto (Activo, Descontinuado, etc.)
- **flag_producto_digital**: Indicador si es producto digital

DIM_CLIENTE

- **cliente_id** (PK): Identificador único de cliente
- **nombre**: Nombre del cliente
- **apellido**: Apellido del cliente
- **email**: Correo electrónico (hash para protección)
- **telefono**: Teléfono de contacto (hash para protección)
- **fecha_nacimiento**: Fecha de nacimiento
- **genero**: Género
- **direccion_id**: ID de la dirección principal
- **ciudad**: Ciudad de residencia
- **region**: Región/Estado/Provincia
- **pais**: País
- **codigo_postal**: Código postal
- **segmento**: Segmento de cliente (Nuevo, Recurrente, Premium, etc.)
- **fecha_primera_compra**: Fecha de primera compra
- **canal_adquisicion**: Canal por el que fue adquirido el cliente
- **lifetime_value**: Valor de vida del cliente
- **preferencia_contacto**: Preferencia de contacto (Email, SMS, etc.)
- **fecha_ultima_compra**: Fecha de última compra

DIM_TIENDA

- **tienda_id** (PK): Identificador único de tienda
- **nombre_tienda**: Nombre de la tienda
- **tipo_tienda**: Tipo (Física, Online, Marketplace)
- **canal**: Canal específico (Web propia, App, Amazon, etc.)
- **direccion**: Dirección física (si aplica)
- **ciudad**: Ciudad
- **region**: Región/Estado/Provincia
- **pais**: País
- **codigo_postal**: Código postal
- **telefono**: Teléfono de contacto
- **horario**: Horario de atención
- **gerente**: Gerente responsable
- **fecha_apertura**: Fecha de apertura
- **tamaño_m2**: Tamaño en metros cuadrados (si es física)
- **estado**: Estado actual (Activa, Cerrada, En remodelación)

DIM_PROMOCION

- **promocion_id** (PK): Identificador único de promoción
- **nombre_promocion**: Nombre de la promoción
- **descripcion**: Descripción detallada

- **tipo_promocion:** Tipo (Descuento porcentual, 2x1, Regalo, etc.)
- **valor_descuento:** Valor del descuento
- **fecha_inicio:** Fecha de inicio
- **fecha_fin:** Fecha de finalización
- **canal_aplicacion:** Canal donde aplica
- **codigo_promocion:** Código promocional
- **presupuesto:** Presupuesto asignado
- **departamento_responsable:** Departamento que lanzó la promoción
- **campana_id:** ID de la campaña asociada
- **nombre_campana:** Nombre de la campaña asociada

DIM_METODO_PAGO

- **metodo_pago_id** (PK): Identificador único del método de pago
- **nombre_metodo:** Nombre del método (Tarjeta de crédito, PayPal, etc.)
- **categoria:** Categoría (Crédito, Débito, Digital, etc.)
- **proveedor:** Proveedor del servicio de pago
- **cuotas_disponibles:** Indica si permite pago en cuotas
- **flag_internacional:** Indica si es internacional

DIM_EMPLEADO (para ventas en tienda física)

- **empleado_id** (PK): Identificador único del empleado
- **nombre:** Nombre del empleado
- **apellido:** Apellido del empleado
- **cargo:** Cargo o posición
- **departamento:** Departamento
- **tienda_id:** Tienda donde trabaja
- **fecha_contratacion:** Fecha de contratación
- **email:** Correo electrónico corporativo
- **telefono:** Teléfono de contacto
- **supervisor_id:** ID del supervisor

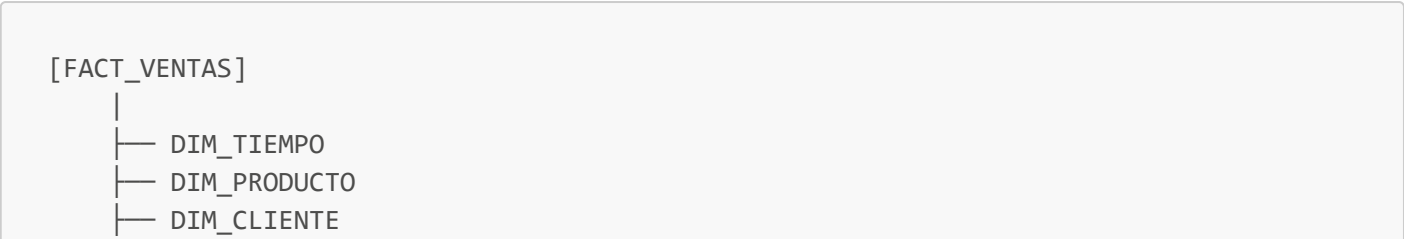
DIM_PAGINA_WEB (para fact_actividad_web)

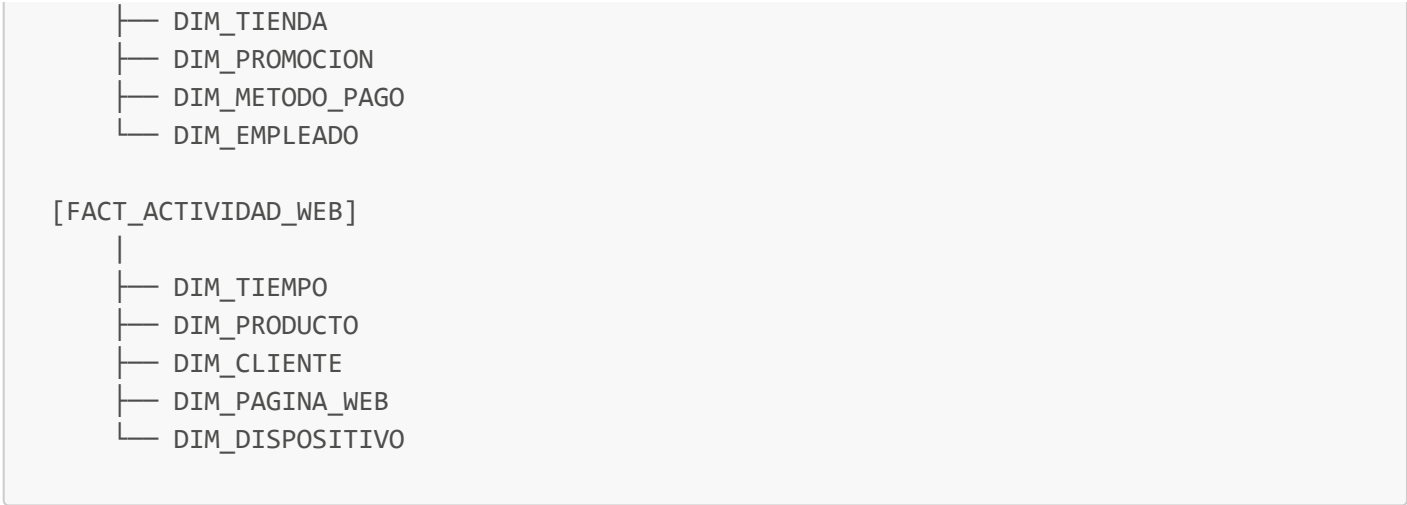
- **pagina_id** (PK): Identificador único de la página
- **url:** URL de la página
- **tipo_pagina:** Tipo (Home, Categoría, Producto, Checkout, etc.)
- **seccion:** Sección del sitio
- **template:** Template utilizado
- **flag_requiere_login:** Indica si requiere login

DIM_DISPOSITIVO (para fact_actividad_web)

- **dispositivo_id** (PK): Identificador único del dispositivo
- **tipo_dispositivo:** Tipo (Desktop, Mobile, Tablet)
- **sistema_operativo:** Sistema operativo
- **navegador:** Navegador utilizado
- **version_navegador:** Versión del navegador
- **resolucion_pantalla:** Resolución de pantalla

3. Diagrama del Modelo





4. Justificación de Decisiones de Diseño

4.1. Elección del Esquema Estrella

Se ha elegido un esquema estrella en lugar de copo de nieve por las siguientes razones:

1. **Simplicidad de consultas:** Reduce el número de joins necesarios para consultas analíticas frecuentes.
2. **Rendimiento optimizado:** Mejor rendimiento para consultas típicas de BI al minimizar las operaciones de join.
3. **Facilidad de comprensión:** Estructura más intuitiva para usuarios de negocio y analistas.
4. **Compatibilidad con herramientas:** Mejor soporte en herramientas de visualización y BI.

El esquema estrella desnormaliza las dimensiones, lo que aumenta el espacio de almacenamiento, pero este trade-off es aceptable considerando las ventajas en rendimiento y usabilidad, especialmente en un entorno moderno con costos de almacenamiento relativamente bajos.

4.2. Decisiones de Normalización/Desnormalización

- **Dimensiones desnormalizadas:** Se ha optado por desnormalizar las jerarquías en dimensiones como Producto (incluyendo categoría, subcategoría y marca) para simplificar las consultas.
- **Dimensiones compartidas:** Dimensiones como Tiempo, Cliente y Producto se comparten entre ambas tablas de hechos para mantener consistencia.
- **Dimensiones role-playing:** La dimensión Tiempo podría implementarse como role-playing para fechas de pedido, envío y entrega si se requiere, pero inicialmente se mantiene una única dimensión para simplicidad.

4.3. Granularidad de Tablas de Hechos

- **FACT_VENTAS:** Granularidad a nivel de línea de pedido (una fila por cada producto en cada transacción), lo que permite análisis detallados por producto.
- **FACT_ACTIVIDAD_WEB:** Granularidad a nivel de actividad (una fila por cada interacción significativa), permitiendo análisis completos del customer journey.

4.4. Métricas Aditivas y No Aditivas

- **Métricas aditivas:** Cantidad, monto_bruto, monto_descuento, monto_impuesto, monto_neto.
- **Métricas semi-aditivas:** precio_unitario, costo_unitario (sumables por producto pero no por tiempo).
- **Métricas no aditivas:** margen (calculada), flags de actividad web.

Para métricas no aditivas, se implementarán cálculos específicos en las herramientas de BI o se precalcularán agregaciones.

4.5. Jerarquías y Agregaciones

Se han identificado varias jerarquías naturales que facilitarán el drill-down/drill-up:

- **Tiempo:** Año → Semestre → Trimestre → Mes → Semana → Día
- **Producto:** Categoría → Subcategoría → Producto

- **Ubicación:** País → Región → Ciudad → Código Postal
- **Cliente:** Segmento → Cliente

Se recomienda la creación de agregaciones precalculadas para las combinaciones más frecuentes:

- Ventas diarias por producto
- Ventas mensuales por categoría y región
- Conversión semanal por canal

5. Integración con la Arquitectura Propuesta

5.1. Fuentes de Datos

Los datos para este modelo dimensional provendrán principalmente de:

- **Zona Trusted del Data Lake:** Datos limpios y validados de transacciones, productos y clientes.
- **Zona Curated del Data Lake:** Datos enriquecidos y con transformaciones avanzadas.

5.2. Flujo de Carga

1. **Extracción:** Datos extraídos de la zona Curated mediante procesos ETL/ELT.
2. **Transformación:** Conversión al modelo dimensional definido.
3. **Carga:** Actualización incremental de dimensiones y hechos en el Data Warehouse.

5.3. Actualización y Mantenimiento

- **Dimensiones:** Proceso SCD (Slowly Changing Dimensions) tipo 2 para dimensiones críticas como Producto y Cliente, manteniendo el historial de cambios.
- **Hechos:** Actualización incremental diaria, con posibilidad de refreshes intradiarios para datos críticos.

5.4. Implementación Técnica

- **Plataforma recomendada:** Snowflake, Redshift o BigQuery
- **Materialización:** Tablas físicas para mejor rendimiento
- **Indexación:** Índices apropiados en claves y columnas de filtrado frecuente
- **Particionamiento:** Por fecha para tablas de hechos

6. Casos de Uso Analíticos

6.1. Análisis de Ventas Multidimensional

```
-- Ventas por categoría, región y mes
SELECT
  p.nombre_categoria,
  t.nombre_region,
  d.nombre_mes,
  d.año,
  SUM(f.cantidad) AS unidades_vendidas,
  SUM(f.monto_neto) AS venta_total,
  AVG(f.margen) AS margen_promedio
FROM
  fact_ventas f
JOIN
  dim_producto p ON f.producto_id = p.producto_id
JOIN
  dim_tienda t ON f.tienda_id = t.tienda_id
JOIN
  dim_tiempo d ON f.fecha_id = d.fecha_id
WHERE
  d.año = 2023
GROUP BY
  p.nombre_categoria,
```

```

t.nombre_region,
d.nombre_mes,
d.año
ORDER BY
d.año, d.mes, venta_total DESC;

```

6.2. Análisis de Conversión y Customer Journey

```

-- Tasa de conversión por tipo de dispositivo y canal
WITH visitas AS (
    SELECT
        dw.tipo_dispositivo,
        t.nombre_tienda AS canal,
        COUNT(DISTINCT aw.sesion_id) AS total_sesiones,
        SUM(aw.flag_carrito_add) AS sesiones_con_add_carrito,
        SUM(aw.flag_checkout_inicio) AS sesiones_con_inicio_checkout,
        SUM(aw.flag_checkout_completo) AS sesiones_con_compra
    FROM
        fact_actividad_web aw
    JOIN
        dim_dispositivo dw ON aw.dispositivo_id = dw.dispositivo_id
    JOIN
        dim_tiempo d ON aw.fecha_id = d.fecha_id
    JOIN
        dim_tienda t ON t.tipo_tienda = 'Online'
    WHERE
        d.mes = 6 AND d.año = 2023
    GROUP BY
        dw.tipo_dispositivo,
        t.nombre_tienda
)
SELECT
    tipo_dispositivo,
    canal,
    total_sesiones,
    sesiones_con_add_carrito,
    sesiones_con_inicio_checkout,
    sesiones_con_compra,
    (sesiones_con_add_carrito / total_sesiones) * 100 AS tasa_add_carrito,
    (sesiones_con_inicio_checkout / sesiones_con_add_carrito) * 100 AS
tasa_inicio_checkout,
    (sesiones_con_compra / sesiones_con_inicio_checkout) * 100 AS
tasa_conversion_checkout,
    (sesiones_con_compra / total_sesiones) * 100 AS tasa_conversion_total
FROM
    visitas
ORDER BY
    tasa_conversion_total DESC;

```

6.3. Análisis de Efectividad de Promociones

```

-- Efecto de promociones en ventas por categoría
SELECT
    p.nombre_categoria,
    pr.nombre_promocion,
    SUM(CASE WHEN f.promocion_id IS NOT NULL THEN f.monto_bruto ELSE 0 END) AS
venta_con_promo,
    SUM(CASE WHEN f.promocion_id IS NOT NULL THEN f.monto_descuento ELSE 0 END) AS
descuento_total,
    SUM(CASE WHEN f.promocion_id IS NOT NULL THEN f.cantidad ELSE 0 END) AS

```

```
unidades_con_promo,  
    AVG(CASE WHEN f.promocion_id IS NOT NULL THEN f.monto_descuento / f.monto_bruto  
ELSE 0 END) * 100 AS porcentaje_descuento_promedio  
FROM  
    fact_ventas f  
JOIN  
    dim_producto p ON f.producto_id = p.producto_id  
JOIN  
    dim_promocion pr ON f.promocion_id = pr.promocion_id  
JOIN  
    dim_tiempo d ON f.fecha_id = d.fecha_id  
WHERE  
    d.año = 2023 AND  
    d.trimestre = 2  
GROUP BY  
    p.nombre_categoria,  
    pr.nombre_promocion  
ORDER BY  
    descuento_total DESC;
```

7. Evolución Futura del Modelo

7.1. Extensiones Planificadas

- **Fact table de devoluciones:** Para análisis de productos con problemas de calidad.
- **Fact table de envíos:** Para análisis de logística y cumplimiento.
- **Dimensión de campañas de marketing:** Para atribución multicanal.

7.2. Consideraciones de Escalabilidad

- **Particionamiento:** Implementar particionamiento por fecha para tablas de hechos de gran volumen.
- **Materialización de agregaciones:** Crear vistas materializadas para consultas frecuentes.
- **Estrategia de purga:** Definir políticas de archivado para datos históricos manteniendo agregaciones.

8. Conclusiones

El modelo multidimensional propuesto proporciona una base sólida para el análisis de ventas y comportamiento de clientes en la empresa de e-commerce y retail digital. La estructura en esquema estrella, con dimensiones compartidas entre múltiples tablas de hechos, facilita el análisis integrado del customer journey desde la navegación web hasta la compra finalizada.

Este diseño se alinea perfectamente con las capas anteriores de la arquitectura de datos, consumiendo datos de calidad de las zonas Trusted y Curated del Data Lake, y preparándolos en un formato optimizado para el análisis de negocio en el Data Warehouse.

El modelo es lo suficientemente flexible para evolucionar junto con las necesidades del negocio, permitiendo la adición de nuevas dimensiones, hechos y métricas a medida que surjan nuevos requisitos analíticos.