

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA DE INGENIERÍA DE SISTEMAS INFORMÁTICOS



CURSO DE ESPECIALIZACIÓN EN INGENIERÍA DE DATOS
TAREA 1 MODULO 3
DESARROLLO DE UN DATA WAREHOUSE TAILSPIN TOYS

PRESENTADO POR:

GONZÁLEZ RIVERA, CÉSAR ALBERTO

CARNET: GR14079

PARA OPTAR AL TÍTULO DE:

INGENIERO(A) DE SISTEMAS INFORMÁTICOS

CIUDAD UNIVERSITARIA, 11 DE AGOSTO DE 2025

Contenido

Introducción a la lógica del negocio.	3
Objetivo General del Proyecto:	3
Modelo Dimensional De Estrella:	3
Descripción del dataset y diccionario de datos del dataset.	4
Origen De Los Datos	4
Tabla 1: Descripción del dataset con su tabla de hechos y dimensiones	6
Descripción del Dataset.	6
Tabla 2: Descripción del dataset con su tabla de hechos y dimensiones	7
Tabla 3: Dataset del sistema OLTP	7
Modelo dimensional propuesto	8
Figura 1: Modelo dimensional propuesto	8

Introducción a la lógica del negocio.

Objetivo General del Proyecto:

Diseñar e implementar un Data Warehouse (DW) para la empresa ficticia Tailspin Toys, con el fin de consolidar y estructurar la información de ventas provenientes de la base de datos transaccional (TailspinToys2020-US).

El DW permitirá realizar análisis históricos y actuales de las ventas en términos de cantidad y valor, considerando cambios en los atributos de productos, estados/regiones y promociones mediante el manejo de dimensiones de tipo SCD2.

Modelo Dimensional De Estrella:

Tabla De Hechos: fact_ventas

Grano: una fila por la línea de pedido (producto por orden).

Claves foráneas: order_date_key de dim_tiempo, ship_date_key de dim_tiempo, product_key de dim_producto, state_key de dim_estado, flags_key de dim_venta_flags

Medidas Principales: quantity cantidad vendida, unit_price – precio unitario aplicado, discount_amount – descuento aplicado, gross_amount – cantidad * precio unitario, net_amount – bruto menos descuento.

Dimensiones:

dim_tiempo

Clave: date_key (YYYYMMDD).

Descripción: almacena el calendario y atributos derivados para análisis temporal.

Campos clave: date, year, quarter, month, daymonth_name, day_name, week_of_year, is_weekend

dim_producto (SCD2)

Clave: product_key (surrogate).

Clave natural: product_id_nk (ProductID en OLTP).

Descripción: almacena datos descriptivos del producto con historial de cambios.

Campos clave: product_sku, product_name, product_category, item_group, kit_type, channels, demographic, retail_price, start_date, end_date, is_current, hash_diff

dim_estado (SCD2)

Clave: state_key (surrogate).

Clave natural: state_id_nk (StateID en OLTP).

Descripción: almacena información del estado y región del cliente.

Campos clave: state_code, state_name, time_zone, region_id, region_name, start_date, end_date, is_current, hash_diff

dim_venta_flags (Junk dimension)

Clave: flags_key (surrogate).

Descripción: combina atributos de promociones y descuentos de baja cardinalidad.

Campos clave: promotion_code, has_discount, is_promo, start_date, end_date, is_current, hash_diff

Descripción del dataset y diccionario de datos del dataset.

Origen De Los Datos

Los datos contenidos en los datasets del Data Warehouse provienen del sistema transaccional de Tailspin Toys, una empresa dedicada a la venta de juguetes, el cual

almacena sus operaciones en una base de datos SQL Server llamada TailspinToys2020-US.

Esta base de datos registra información estructurada sobre productos, clientes, estados/regiones, promociones, transacciones de ventas y otros procesos clave de la operación comercial.

El sistema transaccional está compuesto por diversas tablas relacionales que han sido diseñadas para dar soporte a todas las operaciones diarias del negocio. Dada su complejidad y enfoque operativo, se llevó a cabo un análisis del modelo para identificar las entidades y relaciones relevantes con el objetivo de construir una solución de análisis de ventas.

Se utilizó la metodología de Kimball para diseñar un modelo dimensional orientado al análisis histórico y actual de ventas. El modelo está compuesto por una tabla de hechos y varias tablas de dimensiones, lo que permite realizar análisis detallados desde múltiples perspectivas: temporal, geográfica, de producto y de promociones.

Esta base de datos registra información estructurada sobre:

- Productos y sus atributos comerciales.
- Clientes, incluyendo su ubicación geográfica (estado y región).
- Estados y regiones (zonas de venta).
- Promociones y descuentos aplicados a las ventas.
- Transacciones de ventas, con detalle de cantidades, precios y fechas.

El sistema transaccional está compuesto por varias tablas relacionales diseñadas para soportar las operaciones diarias del negocio. Para el presente proyecto se identificaron como relevantes las siguientes tablas origen:

Tabla Origen (OLTP)	Descripción	Uso en DW
Product	Catálogo maestro de productos (SKU, nombre, categoría, precio de lista, etc.).	Alimenta dim_producto (SCD2).

State	Lista de estados de clientes (código, nombre, zona horaria, ID de región).	Alimenta dim_estado (SCD2).
Region	Catálogo de regiones comerciales.	Complementa dim_estado (SCD2).
Sales	Transacciones de ventas con detalle por producto, cliente, cantidad, precio, descuento y código de promoción.	Alimenta dim_venta_flags (junk dimension) y fact_ventas.
vSalesExtract (vista)	Vista que consolida la información de ventas con los datos necesarios para el DW.	Fuente directa para fact_ventas.

Tabla 1: Descripción del dataset con su tabla de hechos y dimensiones

Descripción del Dataset.

A continuación, se presenta una descripción general del dataset utilizado como base para la construcción del modelo dimensional del sistema de análisis de ventas de Tailspin Toys.

El dataset está compuesto por su tabla de hechos y tablas de dimensiones, organizadas en un esquema en estrella. Cada tabla está estructurada con sus respectivos campos y tipos de datos definidos según los requisitos del modelo dimensional.

En la siguiente tabla se describe el dataset con su tabla de hechos y sus tablas de dimensiones:

DATASET	DESCRIPCIÓN
fact_ventas	Tabla de hechos que registra cada línea de pedido de venta con sus métricas de cantidad, precio y descuentos, asociada a sus dimensiones correspondientes.
dim_producto	Dimensión con información detallada del producto: SKU, nombre, categoría, grupo, tipo de kit, canal, demografía y precio de lista.
dim_estado	Dimensión que describe el estado y región del cliente, incluyendo código, nombre, zona horaria y región comercial.

dim_tiempo	Dimensión temporal para análisis por día, mes, trimestre, año, nombre de día/mes, semana del año y si es fin de semana.
dim_venta_flags	Dimensión “junk” que agrupa atributos de promoción y descuento (código de promoción, indicador de descuento, indicador de promoción).

Tabla 2: Descripción del dataset con su tabla de hechos y dimensiones

A continuación, se describe las tablas asocias del sistema OLTP al DataSet.

DATASET	TABLAS DEL SISTEMA OLTP
fact_ventas	Sales (detalle de ventas), vSalesExtract (vista consolidada de ventas para DW).
dim_producto	Product (catálogo maestro de productos).
dim_estado	State (información de estados), Region (catálogo de regiones).
dim_tiempo	Generada de forma interna en el DW (no depende de tablas transaccionales).
dim_venta_flags	Sales (campos de código de promoción y monto de descuento).
fact_ventas	Sales (detalle de ventas), vSalesExtract (vista consolidada de ventas para DW).

Tabla 3: Dataset del sistema OLTP

Modelo dimensional propuesto



Figura 1: Modelo dimensional propuesto