



**Participante**

Edward Neftalí Liriano Gómez 2022-0437

**Profesor**

Francis Ramírez

**Asignatura**

Electiva 1 (Big Data)

**Tema**

Carga de dimensiones del almacén de datos

# Carga de Dimensiones del Data Warehouse

Sistema de Análisis de Ventas (SDV)

## Introducción

### 1.1 Propósito

Este documento describe el proceso de carga de las tablas de dimensiones del Data Warehouse para el Sistema de Análisis de Ventas (SDV). La implementación sigue los principios de Clean Architecture y utiliza el patrón SCD Tipo 2 (Slowly Changing Dimension) para mantener el historial de cambios.

### 1.2 Alcance

El documento cubre:

- Descripción de cada tabla de dimensión
- Proceso de carga con SCD Tipo 2
- Arquitectura de los Dimension Loaders
- Resultados de ejecución

## Arquitectura del Proceso

### 2.1 Flujo ETL

El proceso ETL se ejecuta en dos fases principales:

Fase	Descripción	Componentes
Fase E	Extracción de datos desde fuentes a Staging	CsvExtractor, ApiExtractor, DatabaseExtractor
Fase L	Carga de datos de Staging a Dimensiones	CustomerLoader, ProductLoader, TimeLoader, StatusLoader

### 2.2 Patrón Strategy para Loaders

Cada cargador de dimensión implementa la interfaz IDimensionLoader:

```
public interface IDimensionLoader
{
    string LoaderName { get; }
    Task<DimensionLoadResult> LoadAsync();
}
```

## Descripción de Tablas de Dimensiones

### 3.1 DimCustomer - Dimensión de Clientes

Almacena información de clientes con soporte para cambios históricos mediante SCD Tipo 2.

Campo	Tipo	Descripción
<b>CustomerKey</b>	BIGINT (PK)	Clave surrogate auto-incremental
CustomerID	INT	Clave de negocio del sistema transaccional
FirstName, LastName	VARCHAR(100)	Nombre y apellido del cliente
Email, Phone	VARCHAR	Información de contacto
City, Country	VARCHAR(100)	Ubicación geográfica
<b>StartDate</b>	DATE	Fecha inicio validez (SCD2)
<b>EndDate</b>	DATE	Fecha fin validez (9999-12-31 = activo)
<b>IsCurrent</b>	BOOLEAN	Flag de versión actual

*Nota:* Los campos resaltados en amarillo son para control de SCD Tipo 2.

### 3.2 DimProduct - Dimensión de Productos

Almacena información de productos con historial de cambios de precios mediante SCD Tipo 2.

Campo	Tipo	Descripción
<b>ProductKey</b>	BIGINT (PK)	Clave surrogate auto-incremental
ProductID	INT	Clave de negocio del producto
ProductName	VARCHAR(200)	Nombre del producto
Category	VARCHAR(100)	Categoría del producto
Price	DECIMAL(10,2)	Precio histórico del producto
<b>StartDate, EndDate</b>	DATE	Período de validez del precio (SCD2)
<b>IsCurrent</b>	BOOLEAN	Flag de precio actual

### 3.3 DimTime - Dimensión de Tiempo

Tabla precalculada para análisis temporal. Se genera mediante el procedimiento almacenado PopulateDimTime(2020, 2030).

Campo	Tipo	Descripción
<b>TimeKey</b>	INT (PK)	Clave en formato YYYYMMDD
FullDate	DATE	Fecha completa
Year, Quarter, Month, Day	INT	Componentes de fecha para análisis
MonthName, DayName	VARCHAR(20)	Nombres de mes y día de la semana
IsWeekend, IsHoliday	BOOLEAN	Flags para análisis de patrones

### 3.4 DimStatus - Dimensión de Estados

Contiene los estados posibles de una orden. Son datos fijos insertados en el script SQL inicial.

StatusKey	StatusName	StatusDescription
1	Pending	Orden creada, pendiente de procesamiento
2	Shipped	Orden enviada al cliente
3	Delivered	Orden entregada exitosamente
4	Cancelled	Orden cancelada

## Proceso de Carga SCD Tipo 2

### 4.1 Algoritmo de Carga

El proceso de carga para DimCustomer y DimProduct sigue esta lógica:

- Leer registros de Staging:** Se obtienen los registros más recientes de cada entidad.
- Buscar registro actual:** Se verifica si existe un registro con IsCurrent = true.
- Comparar campos:** Se detectan cambios en los atributos.
- Si hay cambios:** Cerrar registro anterior (EndDate = ayer, IsCurrent = false).
- Insertar nuevo:** Nueva versión con StartDate = hoy, IsCurrent = true.

### 4.2 Comando de Ejecución

```
dotnet run --project src\SDV.WorkerService\SDV.WorkerService.csproj
```

## Resultados de Ejecución

La ejecución del proceso ETL produce la siguiente salida en consola:

```
SISTEMA DE ANÁLISIS DE VENTAS - ETL COMPLETO

FASE 1: EXTRACCIÓN (E) - STAGING
✓ 5000 clientes extraídos correctamente
✓ 50 productos extraídos correctamente
✓ 100 órdenes extraídas correctamente

FASE 2: CARGA (L) - DIMENSIONES DW
✓ DimStatus verificado: 4 estados disponibles
✓ DimTime verificado: 4018 registros
✓ DimCustomer: 5000 insertados, 0 actualizados
✓ DimProduct: 50 insertados, 0 actualizados

PROCESO ETL COMPLETADO
⌚ TIEMPO TOTAL: 00:00:05.234
```

## Resumen y Estadísticas

Dimensión	Procesados	Insertados	Estado
DimCustomer	5,000	5,000	✓
DimProduct	50	50	✓
DimTime	4,018	(precargado)	✓
DimStatus	4	(fijos)	✓

## Conclusiones

6. Se implementó exitosamente la carga de las 4 dimensiones del Data Warehouse.
7. El patrón SCD Tipo 2 permite mantener historial de cambios en clientes y productos.
8. La arquitectura Clean Architecture facilita la extensibilidad y mantenimiento.
9. El uso del patrón Strategy permite agregar nuevos loaders sin modificar código existente.

## Repositorio

Aquí puede apreciar el enlace y código QR al repositorio de Github que está actualizado al momento de estos avances nuevos:

<https://github.com/Neta2603/SistemaDeVentas/tree/main>

