



**Participante**

Edward Neftalí Liriano Gómez 2022-0437

**Profesor**

Francis Ramírez

**Asignatura**

Electiva 1 (Big Data)

**Tema**

Carga de Facts del almacén de datos

# Carga de Tablas de Hechos del Almacén de Datos

Sistema de Análisis de Ventas - Proceso ETL

## 1. Introducción

Este documento describe el proceso de carga de la tabla de hechos (Fact Table) en el Data Warehouse del Sistema de Análisis de Ventas. La tabla FactSales es el componente central del modelo Star Schema, consolidando las métricas de ventas con referencias a las dimensiones previamente cargadas.

## 2. Objetivo

Implementar la fase final del proceso ETL mediante la carga de la tabla de hechos FactSales, que integra los datos transaccionales de ventas con las dimensiones de tiempo, clientes, productos y estados, permitiendo análisis multidimensional del negocio.

## 3. Proceso de Carga de Facts

### 3.1. Arquitectura del Proceso

El proceso de carga de hechos sigue la arquitectura Clean Architecture implementada en el sistema:

- **Entidad:** FactSales (SDV.Core/Entities/Facts)
- **Repositorio:** FactRepository implementa IFactRepository
- **Loader:** FactSalesLoader implementa IFactLoader
- **Caso de Uso:** LoadFactsUseCase orquesta el proceso completo

### 3.2. Fases del Proceso

#### Fase 1: Limpieza de la Tabla

Antes de cada carga, se ejecuta un TRUNCATE TABLE sobre FactSales, eliminando todos los registros existentes.

#### Fase 2: Extracción y Transformación

El FactSalesLoader obtiene los detalles de órdenes desde StagingOrderDetails y realiza los siguientes pasos de transformación:

- Busca la orden correspondiente en StagingOrders
- Obtiene CustomerKey desde DimCustomer (registros IsCurrent)
- Obtiene ProductKey desde DimProduct (registros IsCurrent)
- Obtiene TimeKey desde DimTime basado en OrderDate
- Obtiene StatusKey desde DimStatus basado en Status
- Calcula TotalPrice = Quantity × UnitPrice

### Fase 3: Carga en Batch

Los registros transformados se insertan en batch en FactSales utilizando AddRangeAsync para optimizar el rendimiento. Esta estrategia reduce significativamente el tiempo de carga comparado con inserciones individuales.

## 4. Estructura de la Tabla FactSales

La tabla de hechos FactSales representa el nivel más granular de análisis: cada línea de detalle de una orden. Su estructura incluye claves foráneas a las cuatro dimensiones y las métricas de negocio.

Campo	Tipo	Descripción
<b>SalesKey</b>	BIGINT (PK)	Clave primaria surrogate
<b>CustomerKey</b>	BIGINT (FK)	Referencia a DimCustomer
<b>ProductKey</b>	BIGINT (FK)	Referencia a DimProduct
<b>TimeKey</b>	INT (FK)	Referencia a DimTime
<b>StatusKey</b>	INT (FK)	Referencia a DimStatus
<b>OrderID</b>	INT	Clave de negocio (trazabilidad)
<b>Quantity</b>	INT	Cantidad de unidades vendidas
<b>UnitPrice</b>	DECIMAL(10,2)	Precio unitario histórico
<b>TotalPrice</b>	DECIMAL(10,2)	Precio total de la línea
<b>LoadDate</b>	DATETIME	Fecha de carga ETL

## 5. Grain y Métricas de Negocio

### 5.1. Grain (Granularidad)

El grain de FactSales es: **Una fila por cada línea de detalle de orden (un producto por orden)**. Este nivel de detalle permite análisis a nivel de producto individual dentro de cada orden, manteniendo la trazabilidad completa del negocio.

### 5.2. Métricas Calculadas

La tabla incluye tres métricas principales para análisis de ventas:

- **Quantity**: Cantidad de unidades vendidas del producto
- **UnitPrice**: Precio unitario en el momento de la venta (histórico)
- **TotalPrice**: Calculado como  $\text{Quantity} \times \text{UnitPrice}$

## 6. Flujo del Proceso ETL Completo

El proceso ETL completo ejecuta tres fases secuenciales:

Fase	Proceso	Salida
Fase 1: Extract	Extracción desde CSV, API y Database	Staging Tables pobladas
Fase 2: Load Dims	Carga de dimensiones con SCD Type 2	4 Dimension Tables cargadas
Fase 3: Load Facts	Limpieza + Transformación + Carga de hechos	FactSales poblada y lista para análisis

## 7. Salida de Consola del Proceso

```
21:04:07 INF] [PROCESO ETL COMPLETADO]
21:04:07 INF] [EXTRACCIÓN (staging)
21:04:07 INF] [  • Clientes:      5000
21:04:07 INF] [  • Productos:     50
21:04:07 INF] [  • Órdenes:       100
21:04:07 INF] [  • Detalles Órdenes: 288
21:04:07 INF] [CARGA DIMENSIONES (Data Warehouse)
21:04:07 INF] [  • Total procesados: 9072
21:04:07 INF] [  • Total insertados: 0
21:04:07 INF] [  • Total actualizados: 0
21:04:07 INF] [CARGA TABLAS DE HECHOS (Data Warehouse)
21:04:07 INF] [  • FactSales insertados: 9
21:04:07 INF] [ TIEMPO TOTAL DE EJECUCIÓN: 00:01:11.7480086
```

## Repository

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>

