

Estracción, Transformación y Carga de datos en Almacén de Gasto en Medicamento



Alejandro Silva Rodríguez

Marta Cuevas Rodríguez

Almacenes De Datos
Universidad de Málaga

Septiembre 2024

Índice

1. Introducción	2
2. Objetivos	2
3. Almacén de Datos de NorthWind	2
3.1. Extracción, Transformación y Carga de Datos	2
4. Almacén de Datos de Gasto de Medicamento en UCI	3
4.1. Rediseño del Almacén	3
4.2. Extracción, Transformación y Carga de Datos	3
5. Dificultades Encontradas	3
6. Conclusión	4
7. Acceso al Repositorio	5

1. Introducción

En el contexto hospitalario actual, el monitoreo y la gestión eficiente de los recursos es una necesidad apremiante, especialmente en unidades como la de Cuidados Intensivos (UCI), donde la administración de medicamentos representa una parte sustancial de los costos. A nivel mundial, el incremento en el costo de los medicamentos y la presión financiera sobre los sistemas de salud han impulsado la búsqueda de soluciones que optimicen el uso de los recursos en entornos críticos. Sin embargo, muchas instituciones hospitalarias carecen de herramientas analíticas específicas para monitorizar y analizar de manera detallada el gasto en fármacos, lo que limita la capacidad de identificar patrones de consumo y optimizar la asignación de presupuestos.

En el informe anterior se detalló el diseño e implementación de un almacén de datos diseñado para analizar el gasto en medicamentos en pacientes ingresados en la UCI en hospitales de EE.UU a partir de la base de datos proporcionada por el MIT [1]. Para abordar este desafío, el proceso de extracción, transformación y carga (ETL) se convierte en un componente clave, ya que permite integrar datos de diferentes fuentes, transformarlos en un formato homogéneo y almacenarlos en el almacén de datos. Este enfoque no solo facilita el análisis eficiente del gasto en medicamentos, sino que también garantiza la calidad y coherencia de los datos utilizados para la toma de decisiones.

2. Objetivos

El objetivo principal de este trabajo es implementar procesos de extracción, transformación y carga (ETL) para alimentar de manera eficiente dos almacenes de datos: el almacén NorthwindDW y un almacén diseñado para analizar el gasto en medicamentos en unidades de cuidados intensivos (UCI). Este propósito se concreta en los siguientes objetivos específicos:

- Diseñar y ejecutar un proceso ETL completo para el almacén de datos NorthwindDW, utilizando herramientas y técnicas previamente estudiadas, con el fin de consolidar conocimientos y garantizar la correcta carga de todas sus tablas.
- Corregir el diseño del almacén de datos, de manera que facilite los procesos ETL.
- Implementar un proceso ETL personalizado para un almacén de datos orientado al análisis del gasto en medicamentos en las UCI.

3. Almacén de Datos de NorthWind

El objetivo de esta parte del proyecto consiste en construir un almacén de datos (NorthwindDW) a partir de la base de datos Northwind, ampliamente utilizada para prácticas educativas y de demostración. Para ello, se desarrolló un proceso de extracción, transformación y carga (ETL) que permite mover los datos desde la base de datos origen hacia el almacén, adaptándolos a su nueva estructura.

3.1. Extracción, Transformación y Carga de Datos

El flujo ETL diseñado se compone de múltiples tareas que se ejecutan de manera secuencial y/o paralela. Estas tareas incluyen la carga de diferentes dimensiones como **Employee**, **Category**, **Product**, entre otras, así como tablas relacionadas con jerarquías geográficas como **Continent**, **Country**, **State**, y **City**. Cada tarea se asegura de transformar los datos según las necesidades del almacén y garantizar su integridad y consistencia.

En la Figura 1, se presenta el flujo de control completo con todas las tareas ejecutadas satisfactoriamente. Este diagrama refleja el correcto funcionamiento del proceso ETL y el éxito en la carga de los datos.

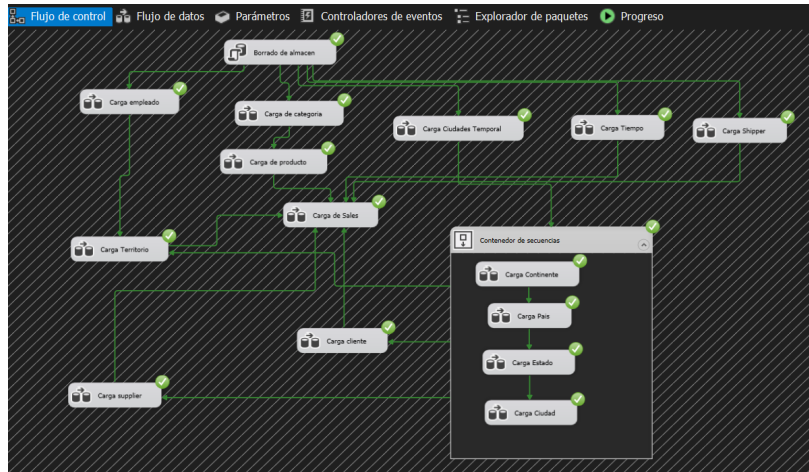


Figura 1: Carga completa del almacén NorthwindDW

Durante la implementación del proceso ETL (Extract, Transform, Load), se encontraron ciertas dificultades. Una de ellas ocurrió durante la carga de la tabla `Employee`. Al intentar insertar datos, surgió el siguiente error relacionado con restricciones de claves foráneas:

```
[Empleado DW [68]] Error: An exception has occurred during
data insertion, the message returned from the provider is:
Se terminó la instrucción. Instrucción INSERT en conflicto
con la restricción FOREIGN KEY SAME TABLE 'FK_Employee_Employee'.
El conflicto ha aparecido en la base de datos 'NorthwindDW',
tabla 'dbo.Employee', column 'EmployeeKey'.
```

Este problema se debía a que la tabla `Employee` contenía una clave foránea que referenciaba a sí misma, lo cual generaba conflictos al insertar los datos en un orden incorrecto. Para resolverlo, se deshabilitaron temporalmente las restricciones de claves foráneas antes de la carga de datos mediante el comando SQL:

```
EXEC sp_MSForEachTable 'ALTER TABLE ? NOCHECK CONSTRAINT ALL';
```

Una vez finalizada la carga de datos, se volvieron a habilitar las restricciones para asegurar la integridad referencial del almacén con el siguiente comando:

```
EXEC sp_MSForEachTable 'ALTER TABLE ? WITH CHECK CHECK CONSTRAINT ALL';
```

Gracias a esta solución, el proceso de carga se completó exitosamente. Este incidente resalta la importancia de planificar adecuadamente la secuencia de carga de datos y manejar las restricciones de integridad de manera controlada.

4. Almacén de Datos de Gasto de Medicamento en UCI

4.1. Rediseño del Almacén

4.2. Extracción, Transformación y Carga de Datos

5. Dificultades Encontradas

A lo largo del proyecto nos hemos encontrado con varios obstáculos que nos han hecho replantear algunos aspectos del diseño del almacén de datos para la gestión hospitalaria:

1. :

A pesar de estas dificultades, cada uno de estos retos nos ha ayudado a entender mejor las particularidades del sector salud y a preparar un diseño más sólido para el almacén de datos.

6. Conclusión

Gracias a la organización del almacén de datos, se optimiza la consulta de información relevante para la toma de decisiones, facilitando la identificación de patrones de gasto asociados a diversas variables, como tratamientos específicos o características de los pacientes. Esta estructura, además, sienta las bases para futuras ampliaciones o análisis más complejos, promoviendo la escalabilidad y adaptabilidad del sistema. En conclusión, el proyecto aporta un modelo sólido y adaptable para la gestión y análisis de datos en el ámbito hospitalario, contribuyendo a una administración más eficiente de los recursos y a una mejora potencial en la atención a los pacientes.

7. Acceso al Repositorio

Toda la información adicional, incluyendo el código fuente y la documentación completa de este proyecto, está disponible en el repositorio de GitHub [2].

Referencias

- [1] MIT Laboratory for Computational Physiology. eICU Collaborative Research Database. <https://eicu-crd.mit.edu/>, 2020. Último acceso: 8 noviembre 2024.
- [2] Alex Silva. Healthcaredatawarehouse. <https://github.com/AlexSilvaa9/HealthcareDataWarehouse>, 2024. Último acceso: 1 octubre 2024.