

# Optimización del Flujo de Información a través de la Planificación de un Almacén de Datos

Miguel Ángel Pérez Solís\* y Andriy Kryvko

Instituto Politécnico Nacional, ESIME-Zacatenco, Av. IPN, S/N, U.P.A.L.M., Edif. Z4, C.P. 07738, Mexico City, Mexico

\*[miguel.ingtpe@gmail.com](mailto:miguel.ingtpe@gmail.com)

**Resumen**— En la actualidad, muchas organizaciones enfrentan dificultades en la gestión y almacenamiento de datos debido a la falta de un sistema de información estructurado. Este trabajo presenta un método sistémico para la planificación de un almacén de datos, proporcionando una solución accesible y escalable para empresas con recursos limitados o sin áreas especializadas en tecnología.

El método propuesto consta de cuatro etapas fundamentales: (1) Reconocimiento del flujo de datos, donde se identifican los procesos que generan información y se estandarizan los datos clave; (2) Diseño del sistema de almacenamiento, en el que se definen los esquemas de base de datos y se elaboran scripts para la migración y verificación de datos; (3) Transformación y carga de datos (ETL), asegurando la homogenización de la información y la implementación de reglas de seguridad e integridad; y (4) Migración de datos, que finaliza el proceso al importar la información al almacén de datos, complementada con un sistema de verificación.

Este enfoque permite sistematizar el flujo de información dentro de un DBMS. La aplicación del método propuesto contribuye a la optimización del almacenamiento, la seguridad de los datos y la eficiencia organizacional, sentando las bases para una gestión de datos más estructurada y confiable.

**Palabras Clave** — Almacén de datos, planificación de sistemas, ETL, DBMS, gestión de datos.

**Abstract**- Nowadays, many organizations face challenges in data management and storage due to the lack of a structured information system. This paper presents a systematic method for data warehouse planning, providing an accessible and scalable solution for companies with limited resources or without specialized IT departments.

The proposed method consists of four fundamental stages: (1) Data flow recognition, where the processes that generate information are identified, and key data is standardized; (2) Storage system design, in which database schemas are defined, and scripts for data migration and verification are developed; (3) Data transformation and loading (ETL), ensuring data homogenization and the implementation of security and integrity rules; and (4) Data migration, which completes the process by importing information into the data warehouse, complemented by a verification system.

This approach enables the systematization of information flow within a DBMS. The applying proposed method contributes to storage optimization, data security, and organizational efficiency, laying the foundation for a more structured and reliable data management system.

**Keywords** — Data warehouse, system planning, ETL, DBMS, data management.

## I. INTRODUCCIÓN

Cuando se habla de analítica de datos en un mundo digital en constante evolución, el tratamiento y administración de datos por medio de herramientas tecnológicas se ha convertido en un factor clave para la toma de decisiones estratégicas, especialmente en el ámbito empresarial. La gestión eficiente de los datos ha impulsado a distintos actores a trabajar en técnicas especializadas para recopilar, mantener y utilizar los datos de manera segura, eficiente y rentable. Por lo cual es necesario orientar estas soluciones sistematizadas que optimicen el acceso y uso de datos que, en muchos casos, permanecen en sistemas locales con acceso limitado y baja capacidad de integración, dificultando su interconexión y análisis. Lo que es notorio en empresas que no son capaces de adaptar un sistema de información a sus procesos y flujo de datos, por lo que archivos electrónicos son utilizados para el almacenamiento de información.

Los sistemas de gestión de bases de datos (DBMS) han demostrado ser esenciales para garantizar la consistencia y disponibilidad de la información. Estos permiten definir, crear y mantener bases de datos de manera local, con el uso de servidores externos o servicios en la nube, este último con un crecimiento significativo en los últimos años [1]. A través del uso de repositorios de datos alineados a metodologías de almacén de datos (DWH), las organizaciones pueden centralizar y estructurar su información, facilitando su análisis con herramientas de visualización de datos para la toma de decisiones [2]. Este enfoque basado en datos es fundamental para la inteligencia de negocios [3] y con ello crear sistemas capaces de concentrar y administrar grandes volúmenes de datos históricos para optimizar estrategias y procesos empresariales.

El presente documento propone la planificación de un sistema de almacén de datos basado en procesos de empresas que no cuentan con un almacén de datos y que generan información en archivos electrónicos con la finalidad de sistematizar el flujo de información en un gestor almacenamiento y consulta en un entorno impulsado por datos.

## II. MÉTODO

En esta sección se describen los pasos a seguir para la planificación de un almacén de datos, con lo que se busca sistematizar el flujo de información para consulta y análisis de manera inmediata, a partir de una revisión de los procesos que generan los datos, identificar como estos se encuentran estructurados y con ello definir el conjunto de tablas que integraran la base de datos con los conocimientos adquiridos basado en la implementación de nuevos sistemas [4], de este modo, el proceso de migración de registros almacenados en archivos electrónicos a un sistema o la nube optimizará el proceso de almacenamiento y consulta.

Para la planificación de un almacén de datos, se diseñó un enfoque basado en cuatro etapas principales: Reconocimiento de la estructura de datos, diseño del sistema de almacenamiento, transformación de datos (ETL) y migración de datos, con la finalidad de sistematizar el flujo de información. A continuación, se describen cada una de estas fases:

### Paso 1: Reconocimiento de la estructura de datos

En esta etapa se identifican los recursos disponibles que generan información relevante para el sistema. En caso de ser necesario, se lleva a cabo un levantamiento de procesos para analizar como se recopilan los datos y en qué formato son almacenados. Por último, se genera el mapeo de los atributos, tipo de dato y extensión de los archivos electrónicos que serán procesados.

### Paso 2: Diseño del sistema de almacenamiento

Una vez reconocido el flujo de datos, se diseña la estructura del almacén de datos de acuerdo a los siguientes puntos:

#### 2.1: Selección del gestor de base de datos

Se define el DBMS a utilizar adecuado en función de los recursos, volumen de almacenamiento, estándares de seguridad y compatibilidad con herramientas de análisis y visualización de datos.

#### 2.2: Definición de la estructura de tablas

Se elabora un diagrama entidad relación (DER) para el diseño de tablas en la base de datos y de ser necesario se optimiza su estructura de acuerdo a las necesidades de información.

### 2.3: Implementación de la Base de Datos

Se procede a la creación de la base de datos y el modelo de datos identificado en la etapa de reconocimiento.

### 2.4: Carga de archivos y consultas

Se genera el método de carga de archivos al sistema en conjunto con las consultas de verificación de datos para mantener su integridad.

### Paso 3: Transformación de datos (ETL)

Se define el proceso ETL (Extract, Transform, Load) el cual permite la conversión de datos a un formato uniforme y sistematizado dentro del almacén de datos.

Se establecen las reglas, métodos y procesos para su aplicación de manera sistémica.

### Paso 4: Migración de datos

Se lleva a cabo el proceso de migración de datos de los archivos originales al sistema de almacenamiento definido. Este paso se complementa con el método de verificación de datos de los métodos establecidos en los pasos 2.4 y 4.1.

#### 4.1: Verificación de datos

Se genera un informe que permita validar la correcta organización e integración de los datos.

Se almacena un registro histórico de los archivos procesados como evidencia y futuras consultas

A continuación, se presentan las etapas de planificación de un almacén de datos, diseñadas para empresas que aún no disponen de un sistema de información y que requieren optimizar su flujo de datos. Esta propuesta ofrece un enfoque inicial estructurado, permitiendo una transición organizada hacia un sistema eficiente de gestión y almacenamiento de información.

## III. DESARROLLO DEL MÉTODO Y DISCUSIÓN

### III.1 Reconocimiento: Procesos y datos estandarizados

El reconocimiento del flujo de datos dentro del sector empresarial es un paso fundamental en la planificación de un almacén de datos. Para ello, se han identificado dos recursos principales: la documentación de procesos en el Sistema de Gestión de Calidad (SGC) [5] que proporciona una visión estructurada del flujo de datos, permitiendo localizar los archivos de almacenamiento y las interacciones entre los distintos procesos organizacionales. Y el segundo, basado en empresas donde no se cuenta con este tipo de documentación, por lo que es necesario implementar un levantamiento de procesos mediante entrevistas con los responsables de cada área.

Levantamiento de procesos: En la Figura 1, se establece una serie de pasos a seguir para estandarizar la información recopilada por los distintos procesos que están relacionados al

flujo de datos, basado en la etapa de identificación de necesidades de un sistema [6], el método se basa en realizar entrevistas con la finalidad de obtener los datos cuantitativos y cualitativos para identificar las interacciones clave entre los datos y las actividades de negocio, facilitando la posterior estructuración del almacén de datos.

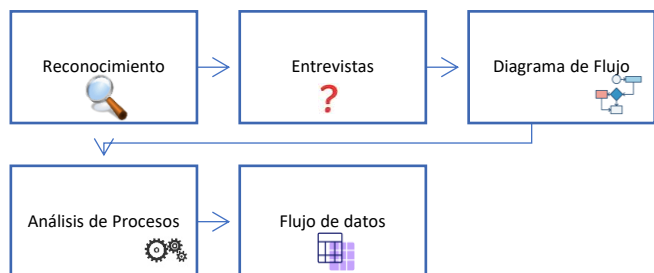


Figura 1. Etapa de identificación de necesidades de un sistema, Metodología del análisis estructurado de sistemas 2001.

Una vez identificado el flujo de datos, se genera el mapeo de los atributos basado en la estructura de las tablas de cada archivo (Encabezado columnas), se definen los tipos de datos (texto, número, fecha etc) y extensión de los archivos electrónicos (xlsx, csv, tx, json, etc) que serán importados al almacén de datos.

### III.2 Diseño del Sistema de Almacenamiento

Los DBMS más utilizados en el sector empresarial varían según el tipo de negocio, la escala de datos y las necesidades específicas de la organización, los sistemas del tipo Relacionales son los más usados en empresas con alta necesidad de integridad y estructura de datos como: Microsoft SQL Server, Oracle Database, MySQL, PostgreSQL. O aquellos servicios en la nube: Amazon RDS, Google Cloud Spanner, Azure SQL Database. Su elección depende de factores como costo, escalabilidad, facilidad de administración y compatibilidad con sistemas existentes. Una vez elegido el DBMS y los productos obtenidos de la etapa de reconocimiento, se elabora el DER con los objetos que dimensionaran la base de datos. Con lo que se busca estructurar el modelo de datos (Etapa 2.3) que contendrá el nuevo sistema y con ello generar los scripts necesarios para el diseño, operación, importación de archivos y consultas de verificación para mantener la integridad de los datos en la BD.

### III.3 Transformación de datos (ETL)

El proceso ETL (Extract, Transform, Load) es una metodología fundamental en la integración de datos previo a su almacenamiento en la base de datos. En esta etapa, se busca homogeneizar los datos provenientes de distintos archivos, asegurando que se ajusten a la estructura y atributos definidos en la base de datos. Este procedimiento es fundamental para

garantizar la consistencia, integridad y calidad de la información almacenada en un DWH. Dado que los procesos de transformación y manipulación pueden alterar los datos, es crucial que las organizaciones establezcan reglas y controles de seguridad que preserven la fiabilidad de la información [7].

La fase de transformación se basa en la normalización de los datos, considerando aspectos como tipos de datos, manejo de valores nulos, eliminación de duplicados y validación de formatos. Estas transformaciones pueden realizarse directamente en los archivos de origen antes de su carga o mediante consultas SQL (queries) dentro del sistema gestor de bases de datos, dependiendo de los requerimientos del sistema y la complejidad de los datos a procesar.

### III.4 Migración de datos

La migración de datos al sistema constituye la etapa final del proceso, junto con la verificación de la información. En esta fase, es fundamental preparar adecuadamente los archivos fuente, asegurando que los datos se ajusten al flujo de información establecido. Esto puede implicar la eliminación de atributos innecesarios en las tablas para optimizar el espacio de almacenamiento y mejorar el rendimiento del sistema.

Para garantizar el esquema de trabajo y facilitar futuras modificaciones o integraciones, el proceso de migración debe ser documentado de manera detallada. Además, los archivos procesados deben ser almacenados en un repositorio seguro como evidencia del contenido cargado. Como parte del control de calidad, se debe generar un informe de verificación, el cual debe incluir un resumen de la información contenida en cada archivo. Este informe, en conjunto con los scripts de consulta del DBMS, permitirá cotejar y validar que los datos migrados coincidan fielmente con la información original.

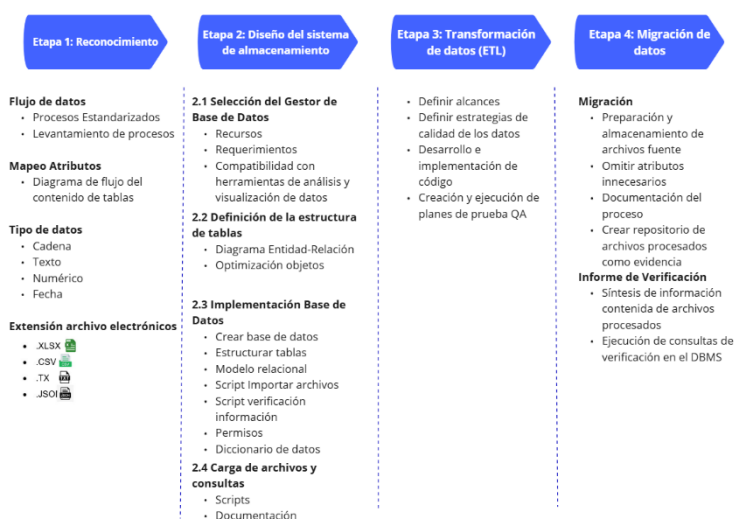


Figura 2. Procesos planificación de un almacén de datos enfocado a la optimización del flujo de datos.

Las distintas etapas abordadas en este método permiten optimizar la gestión de datos, migrando información clave a un almacén de datos indicando la importancia de implementar reglas y controles de seguridad que garanticen su integridad. La Figura 2 presenta el diagrama de flujo de la planificación del almacén de datos, diseñado para sistematizar el flujo de información dentro de un DBMS, lo que facilita una administración más eficiente, segura y estructurada de los datos empresariales.

#### IV. CONCLUSIONES

El método desarrollado proporciona un punto de partida estructurado y accesible para organizaciones que aún no cuentan con un sistema de información definido para el almacenamiento y gestión de datos. Esto suele deberse a limitaciones de recursos o a la ausencia de áreas especializadas en tecnología.

En la Etapa 1, se establecen estrategias para identificar y mapear el flujo de datos, utilizando recursos que detallan las actividades de generación de información, un proceso conocido como estandarización de procesos. Esto permite reconocer y clasificar los datos que serán respaldados en el sistema, así como dimensionar la estructura de las tablas dentro del DBMS.

Los productos obtenidos en esta fase sirven de base para la Etapa 2, centrada en el diseño del sistema de almacenamiento. En esta fase, se abordan las herramientas disponibles para bases de datos relacionales, la definición de esquemas mediante diagramas entidad-relación y la generación de scripts para la migración y verificación de datos.

En la Etapa 3, se destaca la importancia del proceso ETL (Extract, Transform, Load) para homogenizar los datos antes de su incorporación al sistema. Asimismo, se enfatiza la necesidad de implementar reglas y controles que garanticen la seguridad e integridad de la información.

Por último, la migración de datos representa la fase final del método propuesto. Esta etapa, en conjunto con las anteriores, permite sistematizar el flujo de información, asegurando que los datos sean almacenados de manera eficiente en el almacén de datos. Además, la implementación de un sistema de verificación garantiza que la información migrada mantenga su calidad e integridad, estableciendo así un sistema adecuado para una empresa.

Como trabajo a futuro, la integración de este método con herramientas de análisis y visualización de datos puede ser de valor agregado para las organizaciones, mejorando la toma de decisiones. De esta manera, el sistema no solo actúa como un repositorio seguro, sino que también sienta las bases para la automatización y digitalización empresarial.

#### AGRADECIMIENTOS

Los autores agradecen al Consejo Nacional de Humanidades Ciencia y Tecnología y al Programa Institucional de Formación de Investigadores del Instituto Politécnico Nacional por el apoyo brindado

#### DECLARACIÓN DE CONFLICTO DE INTERESES

Los autores manifiestan que no tienen ningún conflicto de intereses financiero, comercial, personal o institucional que pueda haber influido en los resultados o interpretaciones presentadas en este artículo.

#### REFERENCIAS

- [1] J. O. e Sá, R. Gonçalves y C. Kaldeich, «Benchmark of Market Cloud Data Warehouse Technologies» *Procedia Computer Science*, vol. 239, pp. 1212-1219, 2024. <https://doi.org/10.1016/j.procs.2024.06.289>
- [2] A. Khang, R. Gujrati, H. Uygun, R. K. Tailor y S. Gaur, *Data-Driven Modelling and Predictive Analytics in Business and Finance: Concepts, Designs, Technologies, and Applications*, Florida: CRC Press, 2024.
- [3] G. E. Peñafiel, V. M. Zapata Yañes, K. P. Morales Guaman y L. M. Toaquiza Padilla, «Análisis de metodologías para desarrollar Data Warehouse aplicado a la toma de decisiones» *Ciencia Digital*, vol. 3, n° 3.4, pp. 397-418, 2019. <https://doi.org/10.33262/cienciadigital.v3i3.4.922>
- [4] I. Rus, «Model for the integration of software components into heterogeneous IT systems» *Procedia Economics and Finance*, vol. 15, pp. 1098-1104, 2014. [https://doi.org/10.1016/S2212-5671\(14\)00562-0](https://doi.org/10.1016/S2212-5671(14)00562-0)
- [5] «ISO ORG,» [En línea]. Available: <https://www.iso.org/obp/ui/#iso:std:iso:9001:ed-5:v1:es>. [Último acceso: 26 02 2025].
- [6] J. Barranco de Arabea, *Metodología del análisis estructurado de sistemas*, Madrid: Univ Pontifica Comillas, 2001.
- [7] A. Nambiar y D. Mundra, «An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management» *Big Data and Cognitive Computing*, vol. 6, p. 132, 2022. <https://doi.org/10.3390/bdcc6040132>