



**Equipo:** Hernández Flores Angel

Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

## Introducción

El estudio de las bases de datos generalmente inicia con las bases de datos relacionales, son las que principalmente se implementan en las empresas u organizaciones. Las BD relacionales permite almacenar datos y procesar la información generada en base a la operatividad diaria dentro de la misma, por ello se hace hincapié que están diseñadas para la realización de operaciones de consultas y una actualización de los datos de manera eficiente por parte de los usuarios que la manipulan.

El principal objetivo de un Data Warehouse es optimizar el rendimiento de la información almacenada, que queremos decir con esto, que ayude a extraer datos para realizar un análisis, posteriormente hacer una toma de decisiones más rápida.

Este tipo de bases de datos esta basado en un enfoque más eficaz respecto a las bases de datos convencionales.

Mediante este ensayo plasmaremos los orígenes de datos que alimentan un data Warehouse así como también las herramientas utilizadas para los procesos ETL.



**Equipo:** Hernández Flores Angel

Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

## Desarrollo

### **Orígenes de datos que alimentan a un data Warehouse.**

Los data warehouses habían formado un conjunto de datos repetitivos estructurados que eran filtrados antes de entrar en el data warehouse. En los últimos años, el data warehouse ha evolucionado debido a información contextual que ahora se puede adjuntar a los datos no estructurados y que también puede ser almacenada.

Aquellos primeros datos relacionales estructurados no podían ser mezclados y emparejados para temas analíticos con datos textuales no estructurados. Pero con el advenimiento de la contextualización, estos tipos de análisis ahora sí pueden hacerse de forma naturales y fácil.

En el data warehouse, datos no repetitivos, como los comentarios en una encuesta, correos electrónicos y conversaciones, se tratan de forma diferente a las ocurrencias repetitivas de datos, como el flujo de clics, mediciones o el procesamiento máquina o analógico. Los datos no repetitivos son datos basados en textos que fueron generados por la palabra escrita o hablada, leída y reformateada y, lo que es más importante, ahora puede ser contextualizada. Con el fin de extraer cualquier sentido de los datos no repetitivos para su uso en el Data Warehouse, deben tener el contexto de los datos establecidos.



**Equipo:** Hernández Flores Angel

Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

Según Bill Inmon, quien acuñó el término, el data Warehouse se caracteriza por ser:

**Integrado:** Los datos almacenados tienen que ser integrados en una estructura consistente, lo que conlleva a la eliminación de las inconsistencias existentes entre sistemas de operaciones.

**Temático:** Para facilitar el acceso a los datos y el entendimiento de ellos se suelen organizar por temas.

**Histórico:** Permite almacenar datos en momentos concretos, para realizar análisis de tendencias en esos momentos o periodos de tiempo y poder realizar comparaciones entre esos valores tomados.

**No volátil:** Esto implica que la información almacenada en el data Warehouse es para ser leída pero no modificada. Por lo que la información es permanente.

Otra característica importante de la data Warehouse es que contiene metadatos, es decir, datos sobre los datos. Los metadatos permiten saber la procedencia de la información, su periodicidad de refresco, su fiabilidad, forma de cálculo, etc.

### **Ventajas de un data Warehouse.**

Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.

Facilita la aplicación de técnicas estadísticas de análisis y modelización para encontrar relaciones ocultas entre los datos del almacén.

Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.



**Equipo:** Hernández Flores Angel

Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

Simplifica dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.

Supone una optimización tecnológica y económica en entornos de centro de información, estadística o de generación de informes.

### **Fases del Data Warehouse ETL (extracción, transformación y carga)**

Extracción de datos: Extrae datos de distintas fuentes, incluyendo sistemas de BD, aplicaciones, identifica los datos deseados y los extrae.

Transformación: Consiste en el filtrado, limpieza, depuración, homogeneización y agrupación de la información, se lleva a cabo mediante el uso de reglas o tablas de consulta.

Carga: Se cargan los datos al sistema de destino, en este caso a la data Warehouse.

### **Ejemplos Data Warehouse**

Para minoristas: Un almacén de datos o data Warehouse puede ayudar a identificar las características demográficas de los clientes, identificar los patrones de compra y mejorar las respuestas de correo directo.

Para los bancos: Puede ayudar en la detección de fraude de tarjetas de crédito, ayudar a identificar a los clientes más rentables, y poner de relieve los clientes más fieles.

Para fabricantes: Pueden utilizarla para comparar los costos de cada una de sus líneas de productos en los últimos años, a determinar qué factores producen incrementos y ver que efecto tuvieron estos aumentos en los márgenes globales.

**Equipo:** Hernández Flores Angel

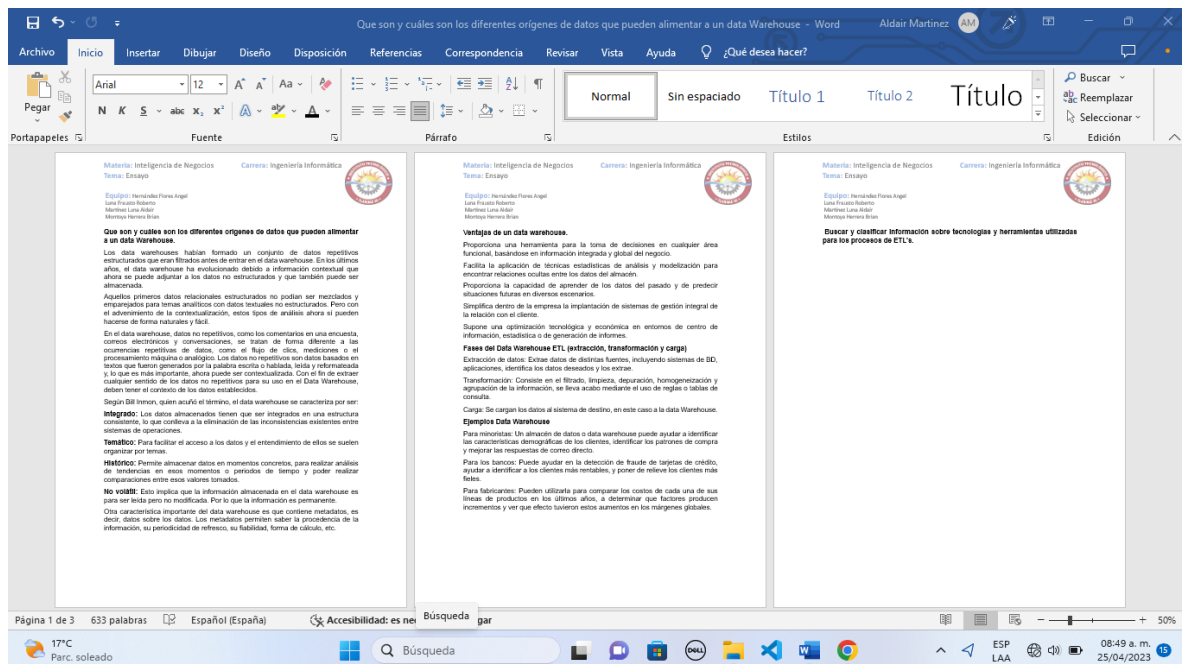
Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

## Tecnologías y herramientas utilizadas para los procesos de ETL's.



Las herramientas ETL son los instrumentos principales que nos permiten construir un data Warehouse o data mart. Sin embargo, no siempre es sencillo saber cómo elegir la herramienta correcta y que mejor se adapte a nuestros objetivos. En este post se explicarán las distintas categorías de herramientas ETL y se hará una comparativa de las principales herramientas en el mercado.



**Equipo:** Hernández Flores Angel

Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

## ¿Qué son las herramientas ETL?

ETL (del inglés extract, transform and load) es el proceso que permite extraer datos de fuentes heterogéneas y con distintos formatos en un único lugar; además, los datos se validan, se limpian y se aplican las transformaciones necesarias para que puedan ser analizados de forma sencilla; finalmente, los datos se cargan en una base de datos, data Warehouse o data mart, donde se encuentran listos para ser explotados, según nuestros objetivos de negocio.

Un proceso ETL puede llegar a ser muy complejo, también teniendo en cuenta el elevado tamaño de los datos para extraer, transformar y cargar. Por lo cual, las herramientas ETL juegan un papel fundamental ya que son la base para cualquier estrategia de análisis de datos y de inteligencia de negocio.



**Equipo:** Hernández Flores Angel

Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

## Categorías de herramientas ETL

Hay distintas herramientas ETL en el mercado, cada una con sus características concretas. Sin embargo, a la hora de elegir la herramienta adecuada para nuestra empresa o proyecto, nos encontramos frente a cuatro distintas categorías principales:

- **Herramientas ETL Enterprise.** Se trata de productos propietarios, con muchas funcionalidades incluidas y soporte para conexión con una gran cantidad de fuentes y suelen ser elegidas por grandes empresas ya que el coste de adquisición es elevado.
- **Herramientas ETL open source.** Se trata de herramientas de código libre y de uso gratuito, lo que permite una mayor accesibilidad para empresas de tamaño reducido. Al ser productos con un enfoque general, es necesaria a menudo una personalización para que se adapten a objetivos concretos, lo que requiere consultoría especializada.
- **Herramientas ETL personalizadas.** Se trata de herramientas desarrolladas a medida y de forma específica para una empresa o proyecto en concreto. Requieren un grande esfuerzo inicial de desarrollo, pero el resultado se ajusta mejor a los requerimientos.
- **Herramientas ETL Cloud.** La nube nos puede proporcionar todas sus ventajas como una alta flexibilidad y el pago por uso a la hora de elegir herramientas ETL que se ofrecen como servicio.



**Equipo:** Hernández Flores Angel

Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

## Características de la herramienta ETL

Antes de realizar una inversión en una plataforma ETL, debe evaluar sus capacidades y características para determinar si cumplirá con sus requisitos de gestión de datos. Estas son algunas características importantes que debe tener una solución ETL:

**Biblioteca de conectores:** Las herramientas ETL modernas ofrecen una amplia biblioteca de conectores, incluidos formatos de archivo, bases de datos y plataformas en la nube. Asegúrese de que la herramienta que compre pueda admitir de forma nativa las fuentes de datos que utiliza.

**Facilidad de uso:** La gestión de asignaciones ETL codificadas a medida es un proceso complejo que requiere una gran experiencia en desarrollo. Para ahorrar recursos de desarrollador y transferir datos de manos de desarrolladores a usuarios comerciales, necesita una solución ETL empresarial que ofrezca un entorno intuitivo y sin código para extraer, transformar y cargar datos.

**Transformaciones de datos:** las necesidades de transformación de datos varían desde trabajos de transformación simples, como búsquedas y uniones, hasta tareas más complejas, como la desnormalización de datos o la conversión de datos no estructurados en tablas estructuradas. Debe seleccionar una herramienta ETL que ofrezca una gama de transformaciones simples y más avanzadas según sus requisitos de manipulación de datos.

**Calidad De Datos Y Caracterización:** solo desea que se carguen datos limpios y precisos en su repositorio de datos. Por lo tanto, busque una plataforma ETL que





**Equipo:** Hernández Flores Angel

Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

ofrezca capacidades integradas de creación de perfiles y calidad de datos para determinar la consistencia, precisión e integridad de los datos empresariales.

**Automatización:** Las grandes empresas deben manejar cientos de trabajos ETL todos los días, lo que solo es posible mediante la automatización. Busque una solución de automatización ETL avanzada con capacidades de automatización de extremo a extremo, incluida la programación de trabajos y la orquestación de procesos, para optimizar los procesos de gestión de datos.

### **Comparativa de herramientas ETL**

Si nos preguntamos cuáles herramientas ETL son mejores, quizás deberíamos antes tener claro para qué nos sirven, ya que cada herramienta tiene sus puntos fuertes que pueden hacer que encaje con nuestro proyecto (y presupuesto). A continuación, vamos a ver las principales características según las cuales es posible realizar una comparativa de herramientas ETL:

El **coste**, que no se limita únicamente el coste de adquisición, sino que también incluye el soporte, la formación y los costes de consultoría. Es importante tener en cuenta el total de estos costes para decidir entre una herramienta propietaria o de código libre.

El **riesgo** de que el proyecto no tenga éxito, lo que incluye no cumplir con el presupuesto, con el calendario o con los requerimientos o expectativas de los clientes.



**Equipo:** Hernández Flores Angel

Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

La **facilidad de uso**, lo que se mejora de forma sustancial si la herramienta dispone de una interfaz gráfica amigable, lo que puede reducir también el tiempo de aprendizaje.

El **soporte y la atención al cliente**. En este sentido hay que tener en cuenta si se ofrece en varios idiomas y países.

Los **requerimientos de despliegue de la herramienta**, lo que incluye la compatibilidad con las distintas plataformas y sistemas operativos, así como los requisitos de sistema en cuanto a hardware.

La **velocidad**, que depende en larga medida de la cantidad de datos que hay que transferir a través de la red y de la capacidad de cálculo requerida para las transformaciones.

La **calidad de datos**, quizás la característica más importante de las herramientas ETL ya que permite disponer de datos validados y limpios.

**Herramientas de control**, que permiten identificar y solucionar los problemas a lo largo de la fase de desarrollo y después.

La **conectividad con todo tipo de sistema**, lo que nos puede permitir extraer datos de todo tipo de aplicaciones heredadas, sean base de datos de Excel, mainframes, ficheros planos, XML, etc.



**Equipo:** Hernández Flores Angel

Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

## Las mejores herramientas ETL

Si queremos saber cuáles son las mejores herramientas ETL, una referencia importante es el Cuadrante Mágico de Gartner, que cada año indica cuáles son los proveedores leader del mercado, entre los cuales se encuentran:

**Informática:** líder según Gartner. Su suite empresarial de integración de datos incluye la solución Information PowerCenter, una de las más populares. Así como también Informática Data Integration.

**IBM:** proporciona la suite de soluciones InfoSphere, en la cual destaca su herramienta DataStage.

**Talend:** conocido por su software de integración de código abierto gratuito Open Studio y por su producto Talend Data Integration.

**SAP:** ofrece la herramienta ETL Data Services como parte de SAP BO (Business Objects)

**SAS:** proporciona una solución de integración de datos llamada Data Management

**Oracle:** proporciona la herramienta Oracle Data Integration, que permite gestionar procesos de integración de datos en sistemas de inteligencia de negocio.



**Equipo:** Hernández Flores Angel

Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

## Conclusiones

### Hernández Flores Ángel

En conclusión a nuestro ensayo hablando de los “Data Warehouse” son una herramienta que se implementa en todas o si no la mayoría de las empresas, principalmente las que manejan una gran cantidad de información, gracias a los data Warehouse pueden tener su información organizada, estructurada y disponer de la información en cualquier momento que sea solicitada, siempre y cuando sea personal autorizado quien la maneje ya que solo ellos pueden tener el acceso, implementar un data Warehouse implica tener una mejor calidad en las bases existentes de la empresa, dejando atrás a las bases de datos lentas y que ocasionalmente había pérdida de información.

Por otra parte, las herramientas “ETL” también son de gran importancia cuando la organización o empresa depende del almacenamiento de sus datos, ya que leen, recopilan y migran grandes volúmenes de información son el primer paso y el más importante en el proceso de almacenamiento de sus datos que permiten la toma de decisiones en el menor tiempo posible.



**Equipo:** Hernández Flores Angel

Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

### **Luna Frausto Roberto**

En conclusión, Las herramientas ETL brindan un lugar central para todos sus datos para que pueda crear canalizaciones de datos y conectar a las partes interesadas y a los líderes con la información que necesitan para trabajar de manera eficiente, también este mismo puede llegar a ser muy complejo ya que teniendo en cuenta el tamaño de los datos en la cuestión de extraer, transformar y cargar datos.

Las herramientas de ETL juegan un papel importante por lo cual nosotros aprendemos a usarlas y entenderlas ya que son fundamentales y sobre todo la base para cualquier estrategia de análisis de datos y de inteligencia de negocios.

Nos enseña que el marco visual ayuda a los usuarios a comprender el flujo de datos desde las fuentes hasta el destino final y estas también vienen con capacidades de monitoreo, manejo de errores y registro para recuperar sus datos de problemas complejos que tenemos día con día en las empresas.

Aprendemos a manejar herramientas de alta importancia con la cual obtenemos nuestra información de manera más compleja y con la capacidad de extraer los datos que la empresa necesita (Lo más útil).



**Equipo:** Hernández Flores Angel

Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

### **Martínez Luna Aldair**

Como conclusión podemos decir que los ETL son proceso que organizan los flujos de datos entre distintos sistemas en una organización y aporta los métodos y herramientas necesarias para movilizar los datos desde múltiples fuentes a un almacén de datos, para transformarlos, limpiarlos y cargarlos en otra base de datos. ETL es parte de la Inteligencia de Negocios, también llamado “Gestión de los Datos” (Data Management).

Analizando lo visto en este ensayo podemos resaltar los siguientes puntos:

Diferentes formas de apoyar el análisis de datos

- Vista tradicional
- ETL;
- almacenamiento ROLAP/MOLAP;
- Optimizaciones lógicas:
  - Vistas materializadas
- Optimizaciones físicas:
  - Índices; fraccionamiento
- OLAP/Data mining para hacer el análisis

**Equipo:** Hernández Flores Angel

Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

.

### **Montoya Herrera Brian**

Como nos hemos dado cuenta a lo largo de este documento las Data Warehouse son tan importantes, así como útiles y no solo en el ámbito empresarial, sino también a nivel microempresa, ya que, gracias a sus fases y procesos, nos ayuda de muchas formas para la organización de datos e información dentro de la empresa, esto debido a que no extrae la información de una sola fuente, si no diversas, dependiendo de la forma en la que la empresa almacena su información, por ejemplo, en bases de datos.

También es muy importante saber identificar que herramienta de ETL es la que mejor nos va para nuestra empresa o proyecto, ya que existen diversas, dichas herramientas son de alta importancia ya que gracias a su proceso de extracción obtenemos la información deseada, en la etapa de transformación se le da forma a la información para obtener solo lo más útil para la empresa, y por último en la etapa de carga es donde la información está lista para ser enviada o subida al personal o clientes de la empresa.



**Equipo:** Hernández Flores Angel

Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

### **Nava Bazan Brenda Vanessa**

Como conclusión podemos decir que los ETL son proceso que organizan los flujos de datos entre distintos sistemas en una organización y aporta los métodos y herramientas necesarias para movilizar los datos desde múltiples fuentes a un almacén de datos, para transformarlos, limpiarlos y cargarlos en otra base de datos.

ETL es parte de la Inteligencia de Negocios, también llamado "Gestión de los Datos" (Data Management).

Las mejores herramientas, Talend, conocido por su software de integración de código abierto gratuito Open Studio y por su producto Talend Data Integration. SAP, ofrece la herramienta ETL Data Services como parte de SAP BO (Business Objects). SAS, proporciona una solución de integración de datos llamada Data Management. Oracle, proporciona la herramienta Oracle Data Integration, que permite gestionar procesos de integración de datos en sistemas de inteligencia de negocio.





**Equipo:** Hernández Flores Angel

Luna Frausto Roberto

**Martínez Luna Aldair**

Montoya Herrera Brian

Nava Bazán Brenda Vanessa

### **Bibliografías:**

- Data Warehouse: todo lo que necesitas saber sobre almacenamiento de datos. (s. f.-a). <https://www.powerdata.es/data-warehouse>
- Rochina, P., & Rochina, P. (2023). ¿Cómo construir un datawarehouse? ¡Te ponemos un ejemplo práctico! Canal Informática y TICS. <https://www.inesem.es/revistadigital/informatica-y-tics/guia-construir-datawarehouse/#:~:text=El%20datawarehouse%20se%20va%20a,clientes%20y%20administraci%C3%B3n%20de%20RRHH.%E2%80%9D>
- Carisio, E. (2019). Herramientas ETL: comparativa y principales categorías. #ADN CLOUD. <https://blog.mdcloud.es/herramientas-etl-comparativa-y-principales-categorias/>
- González, O. (2021). ETL: qué es, procesos y herramientas. *appvizer.es*. <https://www.appvizer.es/revista/it/etl/etl>