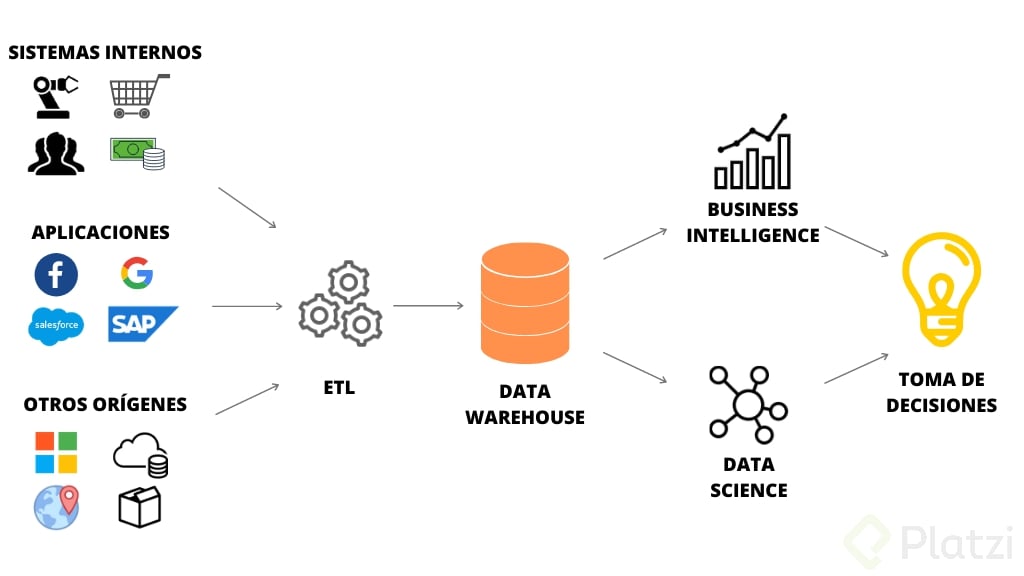
**Data Warehouse**



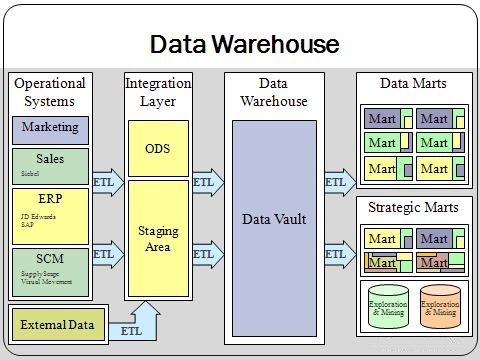
Introducción

Un Data Warehouse es un almacén electrónico donde generalmente una empresa u organización mantiene una gran cantidad de información. Los datos de un Data Warehouse deben almacenarse de forma segura, fiable, fácil de recuperar y fácil de administrar.

Consiste en un repositorio unificado para todos los datos que recogen los diversos sistemas de una empresa. El repositorio puede ser físico o lógico y hace hincapié en la captura de datos de diversas fuentes sobre todo para fines analíticos y de acceso.

Normalmente, un data warehouse se aloja en un servidor corporativo o cada vez más, en la nube. Los datos de diferentes aplicaciones de procesamiento de transacciones Online (OLTP) y otras fuentes se extraen selectivamente para su uso por aplicaciones analíticas y de consultas por usuarios.

Data Warehouse es una arquitectura de almacenamiento de datos que permite a los ejecutivos de negocios organizar, comprender y utilizar sus datos para tomar decisiones estratégicas. Un data warehouse es una arquitectura conocida ya en muchas empresas modernas.



Estructuras de un Data Warehouse

La arquitectura de un data warehouse puede ser dividida en tres estructuras simplificadas: básica, básica con un área de ensayo y básica con área de ensayo y data marts.

* Con una estructura básica, sistemas operativos y archivos planos proporcionan datos en bruto que se almacenan junto con metadatos. Los usuarios finales pueden acceder a ellos para su análisis, generación de informes y minería.
* Al añadir un área de ensayo que se puede colocar entre las fuentes de datos y el almacén, ésta proporciona un lugar donde los datos se pueden limpiar antes de entrar en el almacén. Es posible personalizar la arquitectura del almacén para diferentes grupos dentro de la organización.
* Se puede hacer agregando data marts, que son sistemas diseñados para una línea de negocio en particular. Se pueden tener data marts separados para ventas, inventario y compras, por ejemplo, y los usuarios finales pueden acceder a datos de uno o de todos los data marts del departamento.

Data Warehouse en la nube

Los data warehouse están atravesando actualmente dos transformaciones muy importantes que tienen el potencial de impulsar niveles significativos de innovación empresarial:

* La primera área de transformación es el impulso para aumentar la agilidad general. La gran mayoría de los departamentos de TI están experimentando un rápido aumento de la demanda de datos. Los directivos quieren tener acceso a más y más datos históricos, mientras que, al mismo tiempo, los científicos de datos y los analistas de negocios están explorando formas de introducir nuevos flujos de datos en el almacén para enriquecer el análisis existente, así como impulsar nuevas áreas de análisis. Esta rápida expansión de los volúmenes y fuentes de datos significa que los equipos de TI necesitan invertir más tiempo y esfuerzo asegurando que el rendimiento de las consultas permanezca constante y necesitan proporcionar cada vez más entornos para equipos individuales para validar el valor comercial de los nuevos conjuntos de datos.
* La segunda área de transformación gira en torno a la necesidad de mejorar el control de costes. Existe una creciente necesidad de hacer más con cada vez menos recursos, al mismo tiempo que se garantiza que todos los datos sensibles y estratégicos estén completamente asegurados, a lo largo de todo el ciclo de vida, de la manera más rentable.

La nube está demostrando ser un facilitador dominante. Permite a las organizaciones enfrentarse de forma activa a los desafíos que presentan estas dos transformaciones clave.

La importancia de la implementación de la nube en el uso de las data warehouse está vinculado a tres factores:

* Mayor agilidad: Para mejorar las funcionalidades antes mencionadas, las empresas buscan aprovechar los nuevos flujos de datos y nuevos tipos de análisis que provean de más y mejor información de interés para estas.
* Mejor control de costes: Muchos equipos de TI están buscando maneras de consolidar los data marts existentes, cada uno ejecutándose en hardware dedicado o incluso en hardware propietario, en un único entorno integrado. Los servicios en la nube de calidad ofrecen la oportunidad perfecta para iniciar este tipo de proyectos.

Gracias a la implementación de todo esto, se consiguen 3 ventajas principales al implementar la nube:

* Más fácil consolidación y racionalización.
* Monetización más rápida de los datos en la nube.
* La nube ofrece una protección mayor.

Data Lakes y Data Warehouses

Si bien ambos han estado en auge en los últimos años, se suele confundirlos cuando no son exactamente lo mismo. Podemos decir que, complementan sus utilidades con distintos objetivos.

Estas son las diferencias más significativas entre ambos:

* Datos: En data warehouse los datos son almacenados, mientras que en data lake, no.
* Procesamiento: Antes de que los datos sean almacenados en un data warehouse, estos deben de ser tratados primero, para darles una forma y estructura correcta para su posterior análisis, a diferencia de un data lake, donde los datos se cargan sin este procesamiento y luego, cuando estén listos para usar, se procede con todo este procesamiento de los datos antes mencionados.
* Agilidad: Un almacén de datos es un repositorio altamente estructurado, por definición. No es técnicamente difícil cambiar la estructura, pero puede tomar mucho tiempo dado todos los procesos de negocio que están vinculados a ella. Un data lake, por otro lado, carece de la estructura de un data warehouse, lo que da a los desarrolladores y a los científicos de datos la capacidad de configurar y reconfigurar fácilmente y en tiempo real sus modelos, consultas y aplicaciones.
* Seguridad: La tecnología de la data warehouse ha existido desde hace años, a diferencia de los data lakes, que son más nuevos, por esto mismo, es más factible tener una buena seguridad en un sistema de data warehouse bien desarrollado, que en un data lake.

Data Warehouse - Big Data - BI: diferencias

Los tres conceptos están interconectados y la perspectiva es que, cada vez más, la mayoría de las empresas utilicen el análisis generado por este tipo de tecnologías para tener una visión más analítica de su negocio y así poder tomar las mejores decisiones para crecer.

Se trata de tres conceptos completamente diferentes que tienen en común una nueva manera de lidiar con los datos, siempre teniendo en cuenta la existencia de un gran volumen de información en varios formatos que contribuyen, de forma estructurada o no estructurada, a la toma de decisiones estratégicas. El objetivo final de cualquiera de estas tecnologías es ofrecer una ventaja competitiva a las empresas, pero la forma en que se utiliza es la que marcará la diferencia.

**Big Data**

Llamamos big data a un gran volumen de datos con una variedad, complejidad y velocidad de crecimiento enorme y que además tienen la característica de no ser estructurados. Eso significa que no son relacionales, estando además fuera del entorno corporativo. Es un tipo de tecnología que te permite analizar los datos en tiempo real y puede provenir de diferentes fuentes y formas, tales como mensajería instantánea, redes sociales, registros de grabaciones, imágenes, mensajes de correo electrónico, etc.

**Data Warehouse**

Por otro lado, un data warehouse almacena datos consolidados de diversas fuentes o sistemas de la empresa. Se trata de datos estructurados, que tiene como objetivo principal ser precisos y de alta calidad para de esta forma poder dar soporte a la toma de decisiones de la empresa. Se trata de conseguir tener todos los datos juntos para después poder dividirlos para hacer un análisis de determinados sectores o estrategias.

**Business Intelligence**

Un Business Intelligence (BI) es una especie de “cuello de botella” de los datos recogidos del data warehouse, que llegan de forma exacta y útil para ayudar a la toma de decisiones. Business Intelligence transforma los datos en información útil para analizar no sólo los negocios, sino también las principales estrategias corporativas.