

DATA MART – DATA WAREHOUSE



Universidad
Tecnológica
del Perú

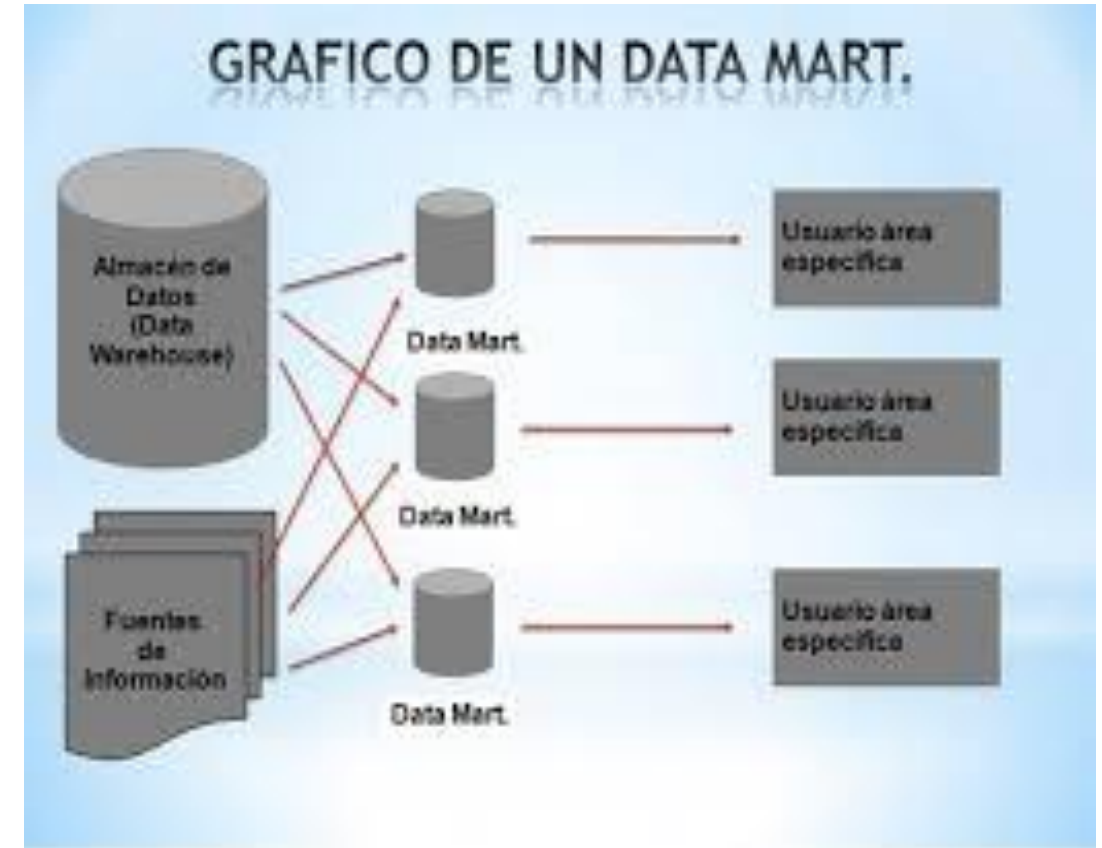
Logro de la sesión

Al finalizar la sesión el estudiante comprende y reconoce que es un data mart y un data warehouse.

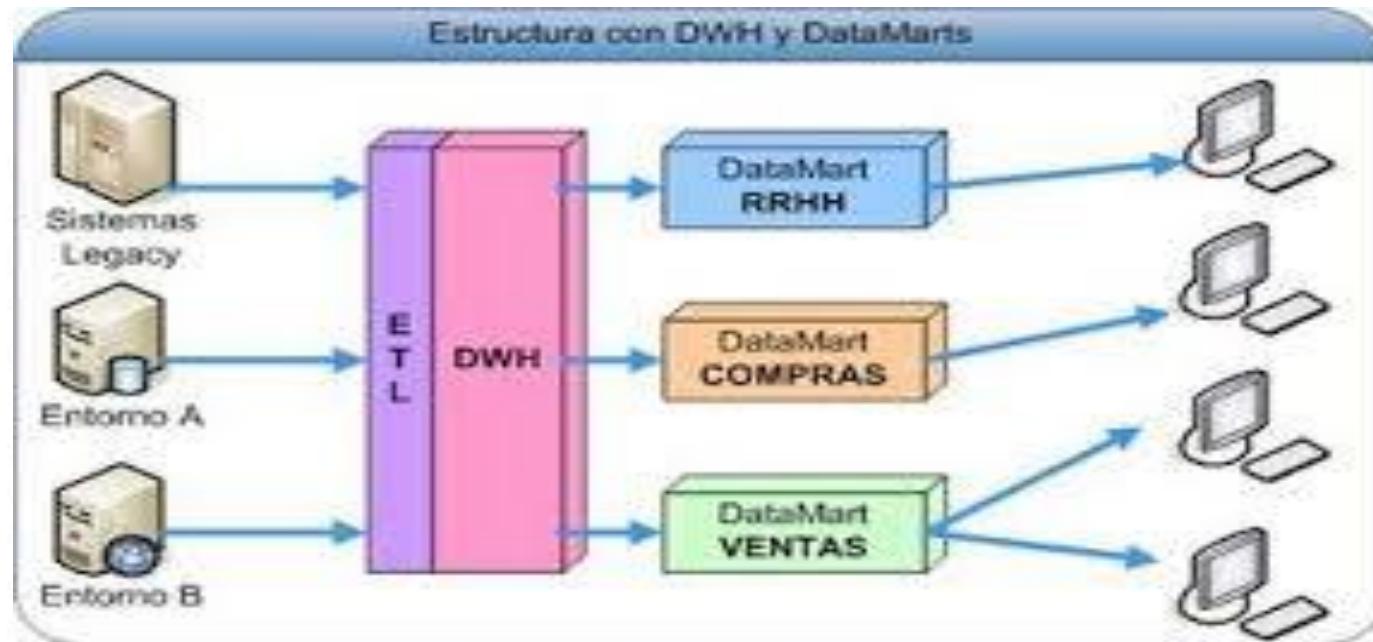


Qué es un Data Mart

Un Datamart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un datamart puede ser alimentado desde los datos de un datawarehouse, o integrar por si mismo un compendio de distintas fuentes de información.



Por tanto, para crear el datamart de un área funcional de la empresa es preciso encontrar la estructura óptima para el análisis de su información, estructura que puede estar montada sobre una base de datos OLTP, como el propio datawarehouse, o sobre una base de datos OLAP. La designación de una u otra dependerá de los datos, los requisitos y las características específicas de cada departamento

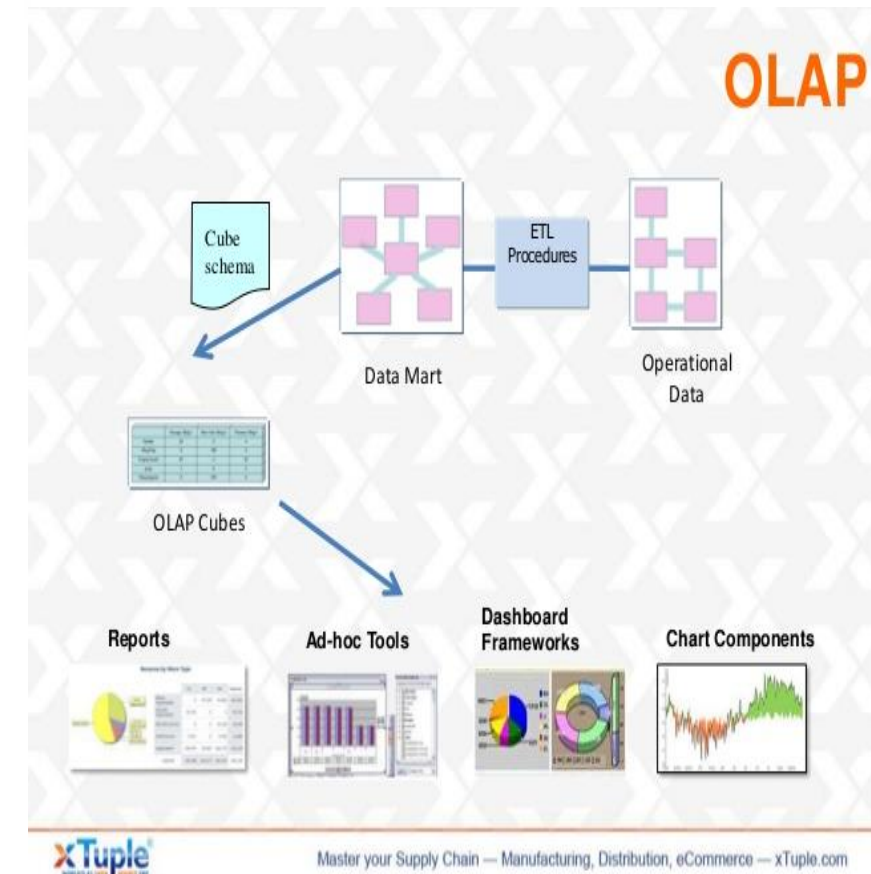


Tipos de Data mart

❑ Datamart OLAP

Los sistemas On-Line Analytical Processing (OLAP) son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejos... etc. Este sistema es típico de los datamarts.

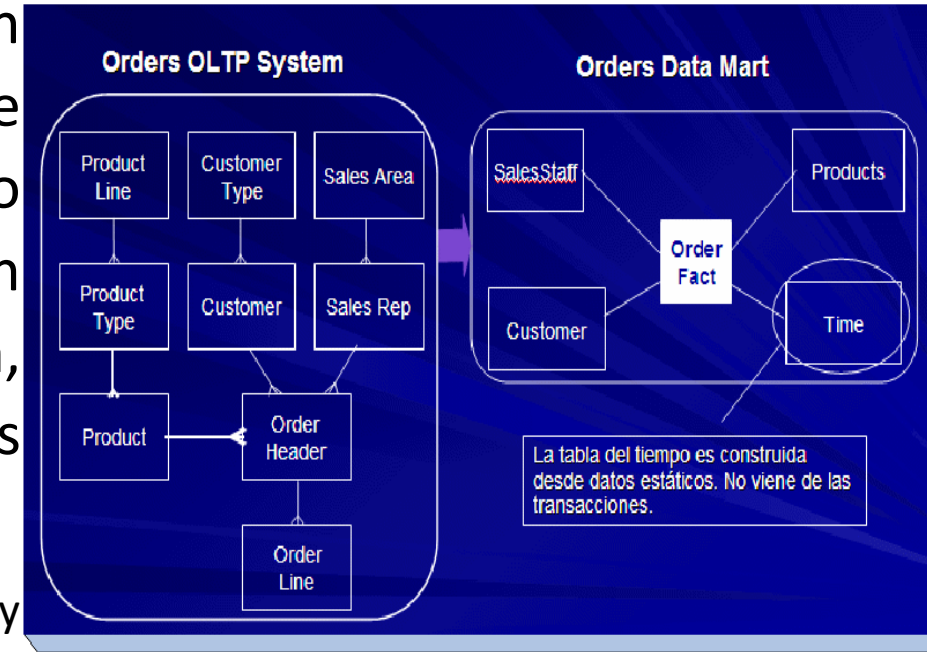
- El acceso a los datos suele ser de sólo lectura. La acción más común es la consulta, con muy pocas inserciones, actualizaciones o eliminaciones.
- Los datos se estructuran según las áreas de negocio, y los formatos de los datos están integrados de manera uniforme en toda la organización.
- El historial de datos es a largo plazo, normalmente de dos a cinco años.
- Las bases de datos OLAP se suelen alimentar de información procedente de los sistemas operacionales existentes, mediante un proceso de extracción, transformación y carga (ETL).



❑ Datamart OLTP

Los sistemas On-Line Transactional Processing (OLTP) son bases de datos orientadas al procesamiento de transacciones. Una transacción genera un proceso atómico (que debe ser validado con un *commit*, o invalidado con un *rollback*), y que puede involucrar operaciones de inserción, modificación y borrado de datos. El proceso transaccional es típico de las bases de datos operacionales.

- El acceso a los datos está optimizado para tareas frecuentes de lectura y escritura. (Por ejemplo, la enorme cantidad de transacciones que tienen que soportar las BD de bancos o hipermercados diariamente).
- Los datos se estructuran según el nivel aplicación (programa de gestión a medida, ERP o CRM implantado, sistema de información departamental...).
- Los formatos de los datos no son necesariamente uniformes en los diferentes departamentos (es común la falta de compatibilidad y la existencia de islas de datos).
- El historial de datos suele limitarse a los datos actuales o recientes.

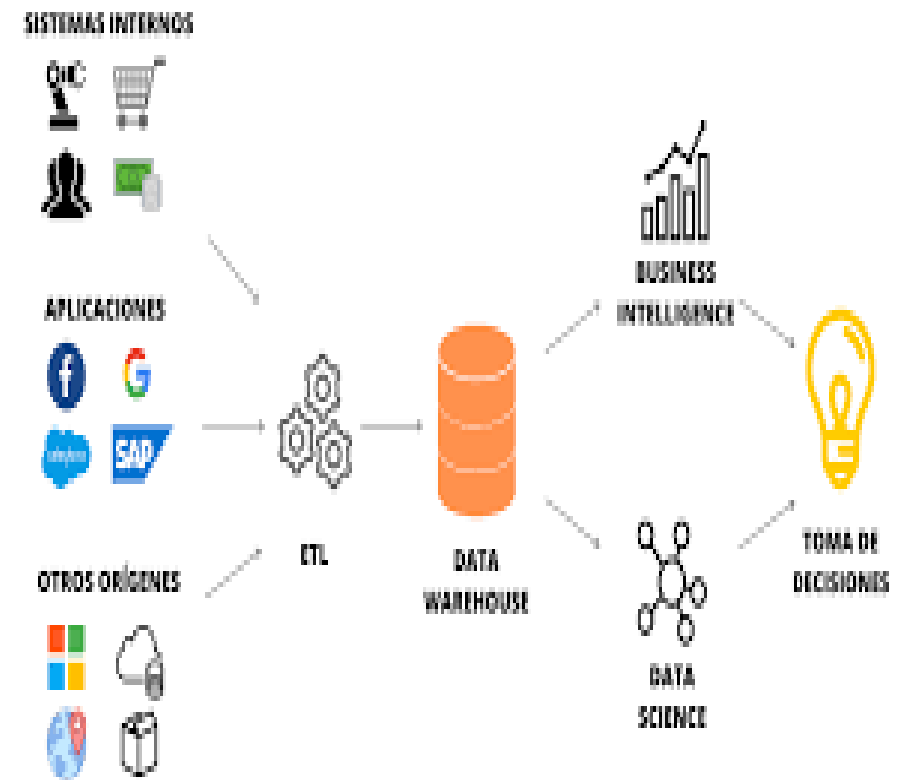


Qué es un Data warehouse

Un data warehouse es 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.

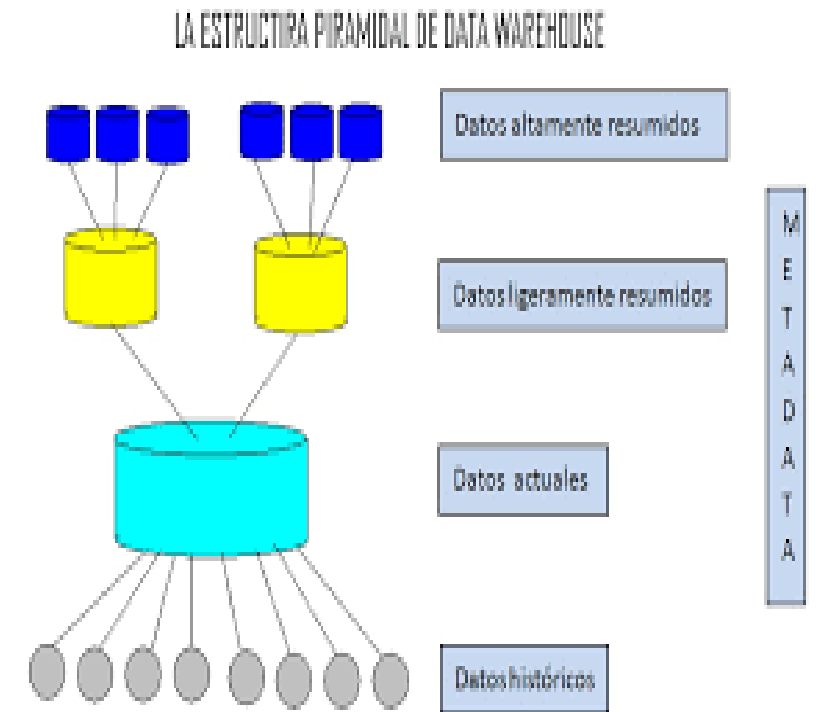
Data Warehouse es una arquitectura de almacenamiento de datos que permite a los ejecutivos de negocios organizar, comprender y utilizar sus datos para tomar decisiones estratégicas. Un data warehouse es una arquitectura conocida ya en muchas empresas modernas.



Estructura 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.



Diferencia entre Data Mart y Data warehouse

Como hemos visto, son términos bastante similares, pero la principal diferencia entre Data Mart y Data Warehouse reside en el alcance.

- Un Data Warehouse es un almacén de datos es un gran depósito de datos recopilados de diferentes organizaciones o departamentos dentro de una corporación.
- Un Data Mart, en cambio, es un subconjunto único de un Data Warehouse. Está diseñado para satisfacer las necesidades de un determinado grupo de usuarios.
- Un Data mart se centra en un solo tema, mientras que los datos de un Data Warehouse comprende datos de todos los departamentos de la organización donde se actualiza continuamente para eliminar datos redundantes
- El proceso de implementación un Data Warehouse puede extenderse muchos meses e incluso años. El proceso de implementación de Data Mart está restringido a unos pocos meses.
- Los datos almacenados en un Data Warehouse siempre ofrecen más detalle en comparación con los de un Data Mart.

Comparativa entre Data Warehouse y Data Mart

Criterio	Data Warehouse	Data Mart
Uso	Ayuda a tomar una decisión estratégica.	Ayuda a tomar decisiones tácticas para el negocio.
Objetivo	El objetivo principal de Data Warehouse es proporcionar un entorno integrado y una imagen coherente de la empresa en un momento determinado.	Utilizado principalmente en una división de negocios a nivel de departamento.
Diseño	Complejo	Sencillo
Modelo dimensional	Puede o no puede usarse en un modelo dimensional. Sin embargo, puede alimentar modelos dimensionales.	Se construye enfocado en un modelo dimensional usando un esquema de inicio.
Gestión de datos	Incluye una gran área de la corporación, por lo que se tarda mucho tiempo en procesarla.	Son fáciles de usar, diseñar e implementar, ya que solo puede manejar pequeñas cantidades de datos.
Foco	El Data Warehouse se enfoca ampliamente en todos los departamentos. Es posible que incluso pueda representar a toda la empresa.	Data Mart está orientado a un área de negocio y se utiliza a nivel de departamento.
Tipo de datos	Los datos almacenados en Data Warehouse siempre ofrecen más detalle en comparación con data mart.	Los Data Marts están contruidos para grupos de usuarios particulares. Por lo tanto, los datos son cortos y limitados.
Normalización	Los almacenes modernos están en su mayoría desnormalizados para proporcionar consultas de datos más rápidas y un buen rendimiento de lectura	No hay preferencia entre una estructura normalizada o desnormalizada.
Área temática	El objetivo principal de Data Warehouse es proporcionar un entorno integrado y una imagen coherente de la empresa en un momento determinado.	En su mayoría tienen solo un área temática, por ejemplo, cifra de ventas.
Almacenamiento de datos	Diseñado para almacenar datos de decisiones de toda la empresa, no solo datos de marketing.	Modelado dimensional y diseño de esquema en estrella empleado para optimizar el rendimiento de la capa de acceso.
Tipo de datos	La variación en el tiempo y el diseño no volátil se aplican estrictamente.	Principalmente incluye estructuras de datos de consolidación para satisfacer las necesidades de informes y consultas del área temática.
Valor de los datos	Solo lectura desde el punto de vista de los usuarios finales.	Datos transaccionales agrupados alimentados directamente desde el Data Warehouse.
Alcance	Es más útil ya que puede traer información de cualquier departamento.	Data Mart contiene datos, de un departamento específico de una empresa. Puede que haya Data Marts separados para ventas, finanzas, marketing, etc. Tiene un uso limitado
Fuente	Los datos provienen de muchas fuentes.	Los datos provienen de muy pocas fuentes.
tamaño	El tamaño del Data Warehouse puede variar de 100 GB a más de un TB.	El tamaño de Data Mart es inferior a 100 GB.
Tiempo de implementación	El proceso de implementación de Data Warehouse puede extenderse de meses a años.	El proceso de implementación de Data Mart está restringido a unos pocos meses.

Conclusiones

En esta sesión hemos aprendido los siguientes conceptos en la Inteligencia de Negocios:

- Qué es un Data Mart
- Datamart OLAP
- Datamart OLTP
- Estructura de un data warehouse
- Diferencia entre Data Mart y Data warehouse
- Comparativa entre Data Warehouse y Data Mart

Fuente: Internet



**Universidad
Tecnológica
del Perú**