

Data Warehouse

Javier Bonet
Joel Catacora

- Base de datos avanzada • 1 de abril del 2015

A large, solid teal circle is centered on a dark, textured background that resembles a scratched or worn surface. The circle is the primary focus of the image.

BUSINESS INTELLIGENCE

El término "Business Intelligence" fue acuñado originalmente por Richard Millar Devens en 'Cyclopædia of Commercial and Business Anecdotes' en 1865. Devens utilizó el término para describir cómo el banquero, Sir Henry Furnese, obtuvo beneficios por recibir y procesar la información sobre su entorno, antes que sus competidores.

La inteligencia de negocios, tal como se entiende, hoy en día se dice que ha evolucionado desde los sistemas de apoyo a las decisiones (DSS), que se inició en la década de 1960 y desarrolló a mediados de los años 80's.

Se denominada **Business intelligence** (BI), al conjunto de estrategias que integran, por un lado el almacenamiento, y por el otro, el procesamiento de grandes cantidades de datos, con el principal objetivo de transformarlos en conocimiento y en decisiones en tiempo real, a través del análisis de los datos existentes en una organización o empresa.

Dato + Análisis = Conocimiento.



¿QUÉ ES UN DATA WAREHOUSE?

Definición en términos de las características del DW:

Data Warehouse (DW o DWH), es una colección de datos orientada a un ámbito determinado, integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza.

Una definición más amplia que la anterior:

Un **Data Warehouse**, es un sistema que extrae, limpia, ajusta, y entrega los datos de origen en un almacén de datos dimensional, y luego apoya e implementa consultas y análisis, con el propósito de la toma de decisiones.



CARACTERÍSTICAS

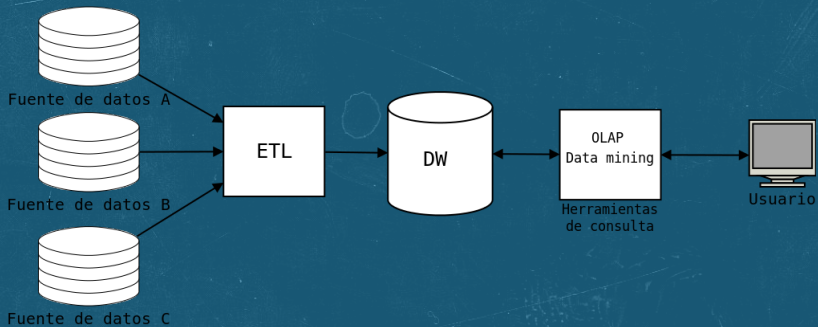
- **Integrado:** Se integran datos provenientes de múltiples fuentes, posiblemente distintas.

- **Integrado:** Se integran datos provenientes de múltiples fuentes, posiblemente distintas.
- **No volátil:** Una vez almacenados los datos en el DW, la información que éstos representan no debe perderse.

- **Integrado:** Se integran datos provenientes de múltiples fuentes, posiblemente distintas.
- **No volátil:** Una vez almacenados los datos en el DW, la información que éstos representan no debe perderse.
- **Variable en el tiempo:** La información histórica se mantiene en el DW a lo largo del tiempo.

A large, solid teal circle is centered on a dark, textured background that resembles a film negative or a scratched surface. The word "ARQUITECTURA" is written in white, uppercase, sans-serif font across the center of the teal circle.

ARQUITECTURA





OLTP VS. DW

	OLTP	Data Warehouse
Objetivo	Soportar actividades transaccionales diarias.	Consultar y analizar información estratégica y táctica.
Modelo de datos	Normalizado.	Desnormalizado.
Datos consultados	Actuales.	Actuales e históricos.
Horizonte de tiempo	60 - 90 días.	5 - 10 años.
Tipos de consultas	Repetitivas, predefinidas	No previsibles, dinámicas
Nivel de almacenamiento	Nivel de detalle.	Nivel de detalle y diferentes niveles de sumarización.
Acciones disponibles	Alta, baja, modificación y consulta.	Carga y consulta.
Número de transacciones	Elevado	Medio o bajo
Tamaño	Pequeño - Mediano.	Grande.
Tiempo de respuesta	Pequeño (segundos - minutos).	Variable (minutos - horas).
Orientación	Orientado a las aplicaciones.	Orientado al negocio.
Sello de tiempo	La clave puede o no tener un elemento de tiempo.	La clave tiene un elemento de tiempo.
Estructura	Generalmente estable.	Generalmente varía de acuerdo a su propia evolución y utilización.

A large teal circle is centered on a dark, textured background that resembles a scratched film or metal surface. The circle is a solid, medium-blue color. Inside the circle, the words "DATA MART" are written in a clean, white, sans-serif font.

DATA MART

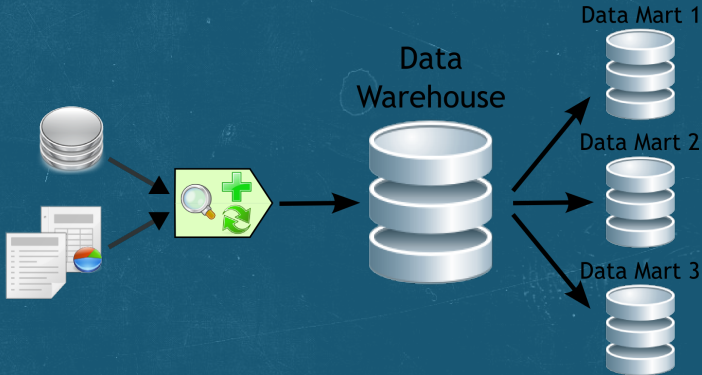
Los **Data marts** (DM), son subconjuntos de datos de un data warehouse para áreas específicas.

Entre las características de un data mart destacan:

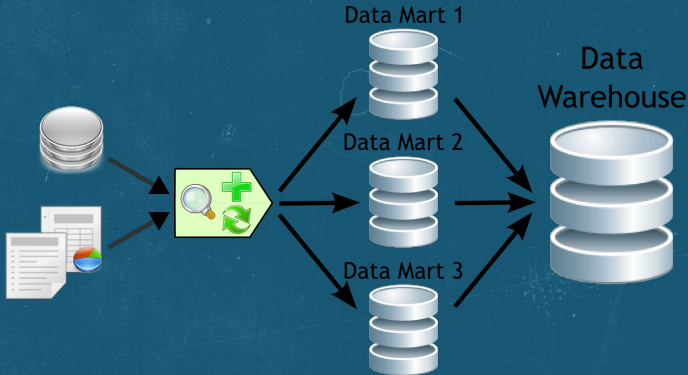
- Usuarios limitados.
- Área específica.
- Tiene un propósito específico.
- Tiene una función de apoyo.

METODOLOGÍAS DE DISEÑO

Top-Down: primero se define el DW y luego se desarrollan, construyen y cargan los DM a partir del mismo.



Bottom-Up: se definen previamente los DM y luego se integran en un DW centralizado.



MODELO DIMENSIONAL

El modelado dimensional es una técnica de diseño lógico de una base de datos, útil para el procesamiento analítico en línea (OLAP), que tiene como ideas centrales el rendimiento y lograr la facilidad de comprensión para el usuario.

Hay dos conceptos básicos, centrales e importantes:

- Hechos (métricas).
- Dimensiones.

ESQUEMA DIMENSIONAL

Sabemos que la relación entre todas las tablas de una base de datos se denomina esquema de base de datos. Para un cierto grupo de bases de datos, en las cuales se realizan consultas sobre datos históricos, generalmente se utilizan diseños llamados esquemas dimensionales.

Un **esquema dimensional** separa físicamente las medidas que cuantifican el negocio (hechos) de los elementos que los describen (dimensiones).

Esquema estrella

En este tipo de esquemas idea central es tener,

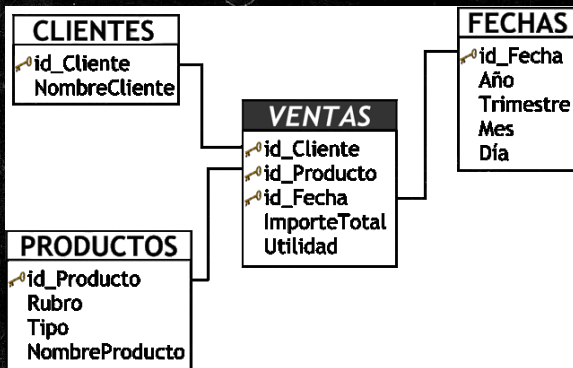
Tabla de hechos

rodeada de

Tablas de dimensiones

Esquema estrella

Supongamos que queremos realizar un análisis sobre el importe ganado por cliente, según el producto y para una fecha específica.



Esquema de copo de nieve

Este otro tipo de esquemas, es similar al esquema de estrella salvo por cierto detalle, las dimensiones pueden, a su vez, estar conectadas con otras tablas de dimensiones.

Tendremos,

Tabla de hechos

rodeada de

Tablas de dimensiones

conectadas con

Nuevas tablas de dimensiones

Esquema de copo de nieve

