

U3. DESARROLLO Y GESTIÓN DE BD DECISIONALES /DATA WAREHOUSING

Clase 2



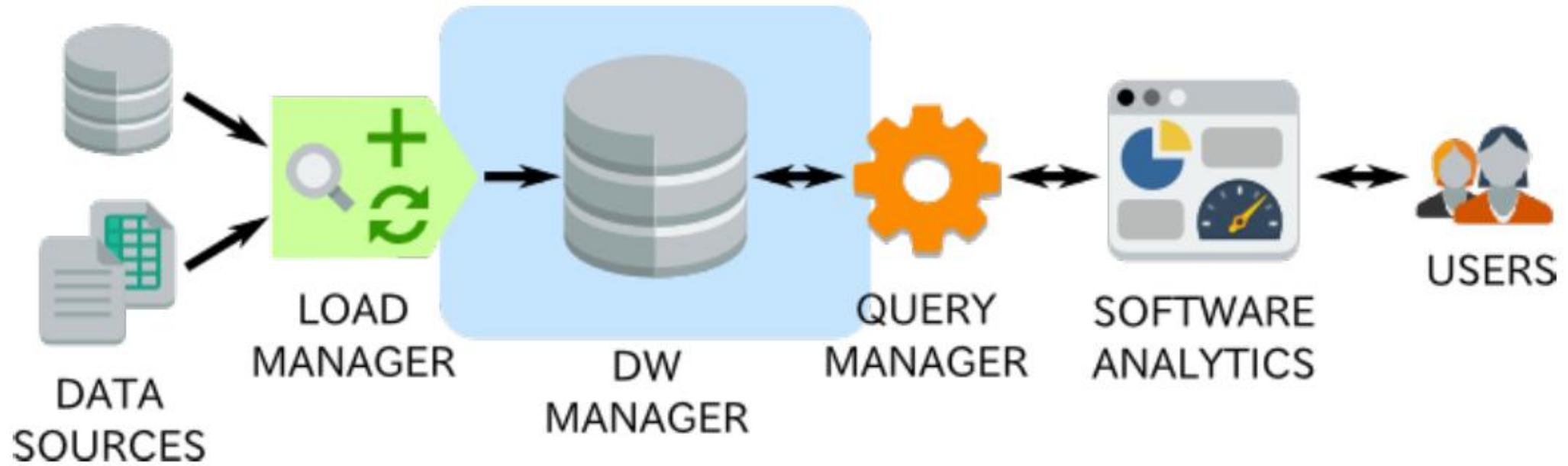
Cátedra: BASE DE DATOS II

Modelo multidimensional: hechos, medidas y dimensiones

Objetivos:

- Profundizar en la arquitectura del DW
- aprender a transformar una necesidad de información en un modelo dimensional.

ESTRUCTURA DEL DW



Su funcionamiento puede resumirse de la siguiente manera:

1. Los datos son extraídos desde distintas fuentes, bases de datos, archivos, servicios web, etc. Estos datos, generalmente, residen en diferentes tipos de sistemas con arquitecturas y formatos variados.
2. Los datos son integrados, transformados y limpiados, para luego ser cargados en el DW. Con base en el DW, se construirán Cubos Multidimensionales y/o Business Models.
3. Los usuarios accederán a los Cubos Multidimensionales, Business Models (u otro tipo de estructura de datos) del DW, utilizando diversas herramientas de consulta, exploración, análisis, reportes, etc.

ESTRUCTURA DEL DW: DATA SOURCES

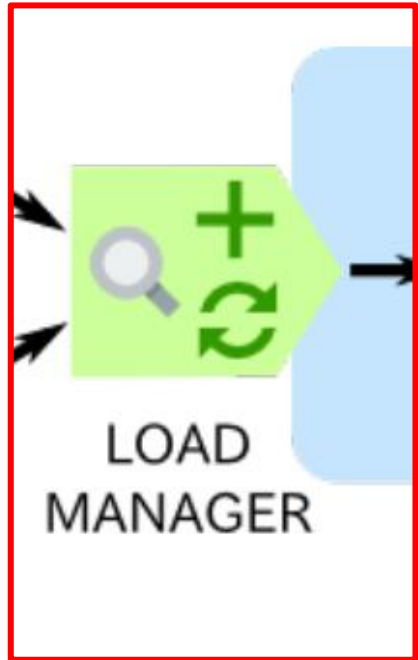


Los Data Sources (orígenes de datos) representan los datos transaccionales que genera la empresa en su accionar diario, junto a otros datos internos y/o externos complementarios. Los Data Sources poseen características muy disímiles entre sí, en formato, procedencia, función, etc.



Archivos de textos, Hojas de cálculos, Informes, Bases de datos transaccionales, SQL y NoSQL, Información no estructurada (páginas web, mails), Redes sociales, RSS, Web Services

ESTRUCTURA DEL DW: LOAD MANAGER



El componente **Load Manager** es el encargado de la ejecución y calendarización (scheduling) de los diferentes procesos de Integración de Datos, siendo el **proceso ETL (Extraction, Transformation, Load)** es la más importante.

- el proceso **Extracción** incluirá técnicas enfocadas por ejemplo a obtener desde diversas fuentes solamente los datos relevantes y mantenerlos en una Staging Area (almacenamiento intermedio);
- el proceso **Transformación** incluirá técnicas encargadas de compatibilizar formatos, filtrar y clasificar datos, relacionar diversas fuentes, etc;
- el proceso **Carga** incluirá técnicas propias de la carga de datos y actualización del DW.



ESTRUCTURA DEL DW: ETL-EXTRACCIÓN

- Una vez seleccionados y extraídos, los datos, deben ser persistidos en una base de datos relacional Staging (almacenamiento intermedio), lo cual permitirá:
 - Manipular los datos sin interrumpir ni sobrecargar los Data Sources y el DW.
 - Crear una capa de abstracción entre la lectura y la carga.
 - Almacenar y gestionar los metadatos que se generan en los procesos ETL.
 - Facilitar la integración.

El Staging Area, generalmente, se constituye en una o más bases de datos relacionales donde la información puede ser persistida en tablas auxiliares, tablas temporales, etc. Una vez que los datos se encuentren en Staging Area, el proceso puede desconectarse de los Data Sources y continuar con la transformaciones necesarias para poblar el DW.

ESTRUCTURA DEL DW: ETL-TRANSFORMACIÓN

Los casos más comunes en los que se debe realizar integración, son los siguientes:

- CODIFICACIÓN**

Grupo etario	Cod-Prestación	Prestación	Dx	Fuente
TODOS	ITQ001W90 - W91	Atención de parto y recién nacido.	Parto	1-Facturación SUMAR
TODOS	ITQ002W88 - W89	Cesárea y atención del recién nacido	Parto	1-Facturación SUMAR
TODOS		Toda prestación con el dx seleccionado	Parto	1-Facturación SUMAR
TODOS	O80 - O84	Egresos	Parto	6-BD Egresos hospitalarios
TODOS		Personas	Parto	8-SIP
TODOS		Personas	Parto	9- Listado de partos del HMyN

- MEDIDA DE ATRIBUTOS**

PESO AL NACER

3100g

Peso: 5.75 kg.

Altura: 61.00 cm.

Masa Corporal: 15.45 kgr/m²

Evaluación nutricional: Adecuado

- FUENTES MÚLTIPLES**

4654	REYNOSO	Ste	21/04/:	Femenin	L00692		21/07/2025	Supervisión embarazo	Z34
4654	REYNOSO REARTE	S	21/04/:	Femenin	L00692		21/07/2025	Supervisión embarazo	Z34

ESTRUCTURA DEL DW: ETL-CARGA

Esta función es responsable de ejecutar las tareas relacionadas con Carga Inicial (Initial Load) y Actualización periódica (Update).

- La Carga Inicial (Initial Load), se refiere a la primera carga de datos que recibe el DW. Generalmente, esta tarea consume un tiempo considerable, debido a que se insertan gran cantidad de registros pertenecientes a períodos largos de tiempo.
- La Actualización Periódica (Update), se refiere a la inserción de pequeños volúmenes de datos, y su frecuencia está dada en función de la granularidad (cuán resumidos se encuentran los datos) del DW y los requerimientos de los usuarios. El objetivo de esta tarea es añadir al DW solo aquellos datos que se generaron a partir de la última actualización (delta de cambios).

ESTRUCTURA DEL DW: DATA WAREHOUSE MANAGER

El DW Manager está compuesto por una serie de aplicaciones de software dedicadas a gestionar:

-  el DW (SGDB),
-  las conexiones a base de datos y otros Data Sources,
-  las estructuras de datos (Cubos Multidimensionales, Business Models, etc.),
-  información de autenticación y autorización (credenciales de acceso, users, roles, permisos, etc), y
-  otros metadatos.



BASE DE DATOS MULTIDIMENSIONAL

Un **Data Warehouse es una base de datos con estructura multidimensional**, lo cual se traduce en una forma específica de almacenamiento en la cual se definen dos estructuras principales:

- tablas de Hechos y
- tablas de Dimensiones.

La utilización de tablas de Hechos y Dimensiones, facilita la creación de estructuras de datos (Cubos Multidimensionales, Business Models, etc.) y posibilita que las consultas al SGBD sean respondidas con mucha performance.

BASE DE DATOS MULTIDIMENSIONAL

- **Hecho:** qué fenómeno se quiere analizar.
- **Medidas:** qué se quiere medir.
- **Dimensiones:** desde qué perspectivas se analiza.
- **Granularidad:** qué representa una fila de la tabla de hechos.
- **Jerarquías** dentro de las dimensiones.
- Ejemplos de dimensiones: **tiempo**, geografía, establecimiento, paciente, profesional.

¿Qué pasó? → ¿Qué quiero medir? → ¿Cómo quiero analizarlo?

En este caso:

¿Qué pasó?

→ Una internación.

→ **HECHO**

¿Qué quiero medir?

→ Cuántas internaciones.

→ **MEDIDA**

¿Cómo quiero analizarlas?

→ Por hospital, diagnóstico, edad y mes.

→ **DIMENSIONES**

BASE DE DATOS MULTIDIMENSIONAL

Ejercicio:

Una compañía dedicada a la venta de muebles necesita analizar información sobre ciertas medidas de su negocio. Para ello se decide diseñar un Almacén de Datos (Data Warehouse) para la empresa.

Se pide:

1. Identificar hechos, medidas y dimensiones.
2. Diseñar un esquema en *estrella* para el almacén de datos, asumiendo algún concepto de *jerarquía* para cada una de las dimensiones. Considerando que se quiere responder a los siguientes requerimientos:
 - a. Obtener la cantidad total de ventas, de ingresos y de descuentos respecto de cada ciudad, cada tipo de mueble y por mes.
 - b. Obtener el promedio de ventas, de ingresos y de descuentos anual respecto de cada región, y categoría de mueble.

BASE DE DATOS MULTIDIMENSIONAL

1. Identificar hechos, medidas y dimensiones.

Hecho —————> Concepto relevante en el proceso de toma de decisiones

➤ Ventas

Medidas —————> Propiedad numérica de un hecho

➤ Cantidad de ventas

➤ Ingresos por ventas

➤ Descuentos por ventas

Dimensiones —————> Propiedad de un hecho respecto de un dominio finito

➤ Muebles (Tipo , Categoría , Material)

➤ Clientes (Nombre, Edad, Sexo , Ciudad, Provincia, Región)

➤ Tiempo (Día, Mes , Año)