

U3. DESARROLLO Y GESTIÓN DE BD DECISIONALES /DATA WAREHOUSING

Clase 1



Cátedra: BASE DE DATOS II

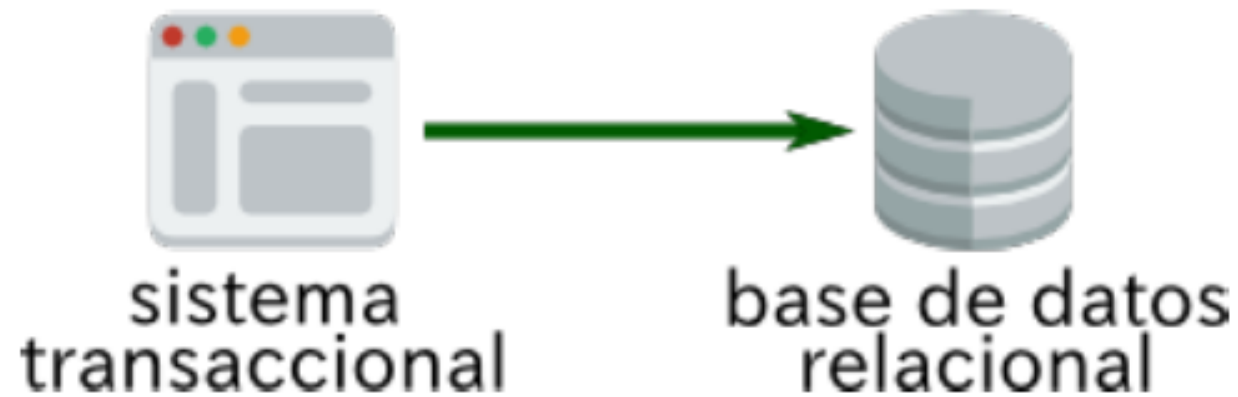
INTRODUCCIÓN A DATA WAREHOUSE

Objetivos:

- Comprender los conceptos básicos de un Data Warehouse.
- Conocer los beneficios y aplicaciones de un Data Warehouse.
- Familiarizarse con la arquitectura de un Data Warehouse.
- Identificar las principales etapas del proceso de construcción de un Data Warehouse.
- Iniciar sobre una necesidad local de requerimientos de datawarehousing

DATOS COMO PRODUCTOS SECUNDARIOS

En toda organización se generan datos constantemente, para la ejecución de sus operaciones y transacciones. Es muy común, que estos datos se administren a través de sistemas transaccionales y se almacenen en bases de datos relacionales, aunque esto no es excluyente.



Al pasar de los años, la acumulación de estos datos NO produce utilidad alguna, o lo hace de manera poco relevante a través de esporádicas consultas históricas.

¿Qué pasaría si de alguna manera procesamos todos estos datos y los utilizamos como fuente de información para la toma de decisiones?

¿Qué pasaría si además contamos con herramientas de software especializadas en la presentación de los datos para el estudio analítico?

TÉRMINOS CLAVES

- **OLTP** o procesamiento de transacciones en línea es un tipo de procesamiento de datos que consiste en ejecutar una serie de transacciones que ocurren simultáneamente. En la banca en línea, las compras, la entrada de pedidos o el envío de mensajes de texto, por ejemplo.
- **Datos transaccionales** es información que se captura de las transacciones.
Ejemplificando en un área comercial: registran la hora de la transacción, el lugar donde se realizó, los artículos comprados, y cualidades asociadas con la transacción.
- **Fuentes:** sistemas, archivos y otros que alimentan al DW. Pueden ser externas o internas.

DEFINIENDO AL BI

Se puede describir el **Business Intelligence (BI - Inteligencia de Negocios)**, como la actividad de:

- almacenar y procesar grandes cantidades de datos,
- para que mediante la utilización de herramientas de software especializadas,
- sea sencillo el análisis y exploración de dichos datos,
- con el principal objetivo de obtener conocimiento (knowledge) orientado a tomar decisiones en tiempo real.

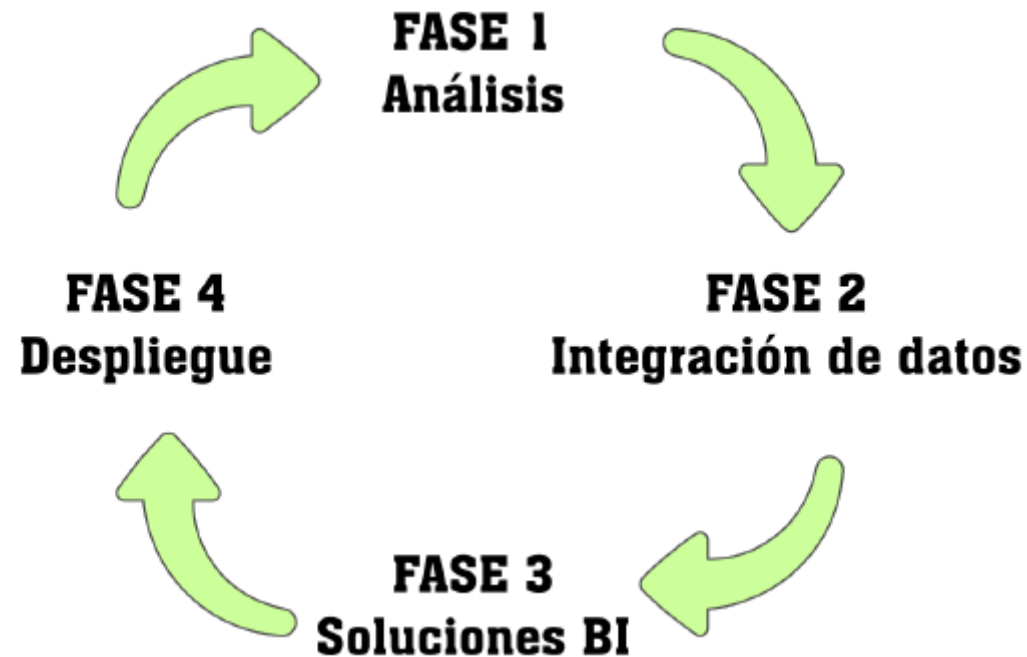


- lo que está sucediendo en la organización,
- lo que ha sucedido,
- lo que puede llegar a suceder y
- por qué.

PROCESO BI

FASE 1: Análisis. El foco estará puesto en conocer las necesidades de información de la organización, a partir de las cuales se redactarán las preguntas.

FASE 2: Integración de datos. Se extraen e integran datos de diferentes fuentes (data sources), de origen interno y/o externo. Para la integración se les aplicará transformaciones a fin de compatibilizarlos con los requerimientos del análisis, y posteriormente se cargarán en la base de datos final, cuyo propósito es servir a BI.



FASE 4: Despliegue. En esta fase se les entregará a l@s usuari@s los componentes BI que le correspondan y se les capacitará en su utilización, como así también en la creación de sus propias soluciones BI.

FASE 3: Soluciones BI. Herramientas, técnicas y componentes que permitirán la explotación de los datos. Algunos más utilizados son: reportes, indicadores, análisis interactivos, dashboards, gráficos estadísticos, etc.

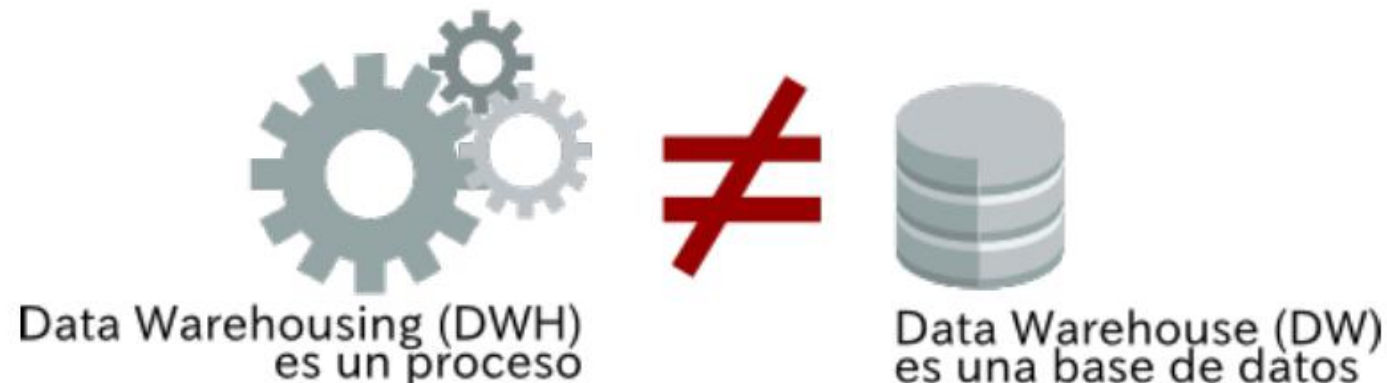
DATA WAREHOUSING

El **Data Warehousing** (DWH), es el proceso que reúne y ordena las tareas inherentes a:

- la extracción, transformación, consolidación, integración y
- centralización de los datos internos (datos que una organización genera en su actividad diaria, por ejemplo compras, ventas, producción, etc-) y los datos externos relacionados;
- permitiendo de esta manera el acceso, análisis y exploración,
- con el objetivo de dar soporte al proceso de toma de decisiones estratégico y táctico.

Para almacenar los datos necesarios para el análisis, se empleará un Data Warehouse (DW).

Un **Data Warehouse** es una base de datos que posee una estructura multidimensional.



¿QUÉ ES UN DATA WAREHOUSE?

Fue acuñado por primera vez por Bill Inmon en 1990.

Conjunto de datos estructurados orientados por temas integrados, variables con el tiempo y no volátiles, empleados para tomar decisiones.

- Un Data Warehouse, es una arquitectura de almacenamiento diseñada para contener datos extraídos de sistemas de transacciones, almacenes de datos operativos y fuentes externas. Luego, el almacén combina esos datos en una forma de resumen agregada adecuada para el análisis de datos en toda la empresa y la generación de informes para necesidades comerciales predefinidas.
- **Extraer los datos de los sistemas transaccionales -> Reestructurarlos (forma unificada, estándar)-> Consumidos por BI -> Toma de Decisiones**

FUNDAMENTOS DE UN DATA WAREHOUSE





ORIENTADO A OBJETOS / AL NEGOCIO

- al DW solo ingresarán **datos relevantes para el análisis y toma de decisiones**. Por ejemplo, los datos referidos a l@s client@s, como su dirección de correo electrónico, fax, teléfono, código de identidad, código postal, etc, que son tan importantes de almacenar en cualquier sistema transaccional, NO son tenidos en cuenta para el DW por carecer de valor analítico, pero sí lo son aquellos que indican el tipo de cliente, su clasificación, ubicación geográfica, edad, etc.
- se manejarán **entidades de alto nivel**, es decir si una empresa maneja stock, listas de precios, cuentas corrientes, ventas, compras, etc., con lo que trabajaremos en el DW serán con entidades del tipo: clientes, productos, rubros, zonas, etc.
- **la estructura será multidimensional**, es decir que almacenará sus datos en tablas de Hechos y tablas de Dimensión.



INTEGRADO

- La integración implica que todos los datos provenientes de orígenes heterogéneos deben ser analizados a fin de asegurar su calidad y limpieza para luego ser consolidados en el DW.
- El proceso que permite esta consolidación, se denomina Integración de Datos, y cuenta con diversas técnicas y subprocesos para llevar a cabo sus tareas. Una de estas técnicas es el proceso **ETL: Extracción, Transformación y Carga de datos** (Extraction, Transformation and Load).



No Volátil

NO VOLÁTIL



DATA SOURCES



Consultar



Insertar



Eliminar



Modificar



DATA WAREHOUSE



Consultar



Insertar

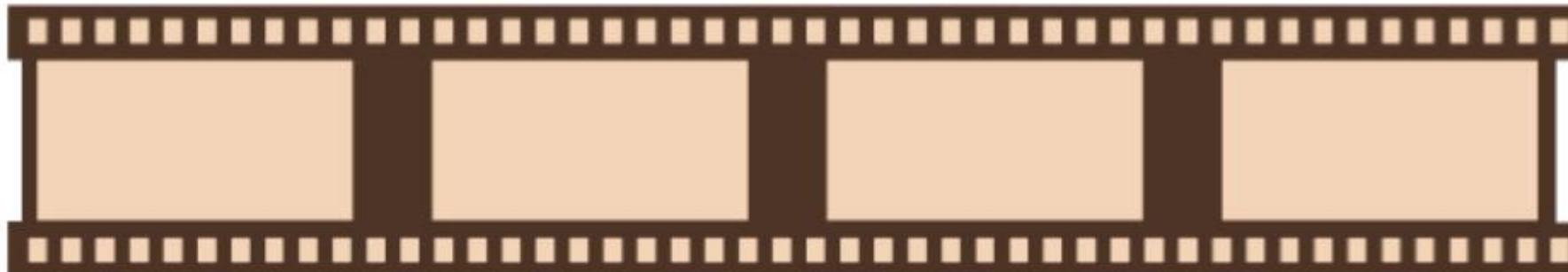


Variante en
el Tiempo

VARIANTE EN EL TIEMPO

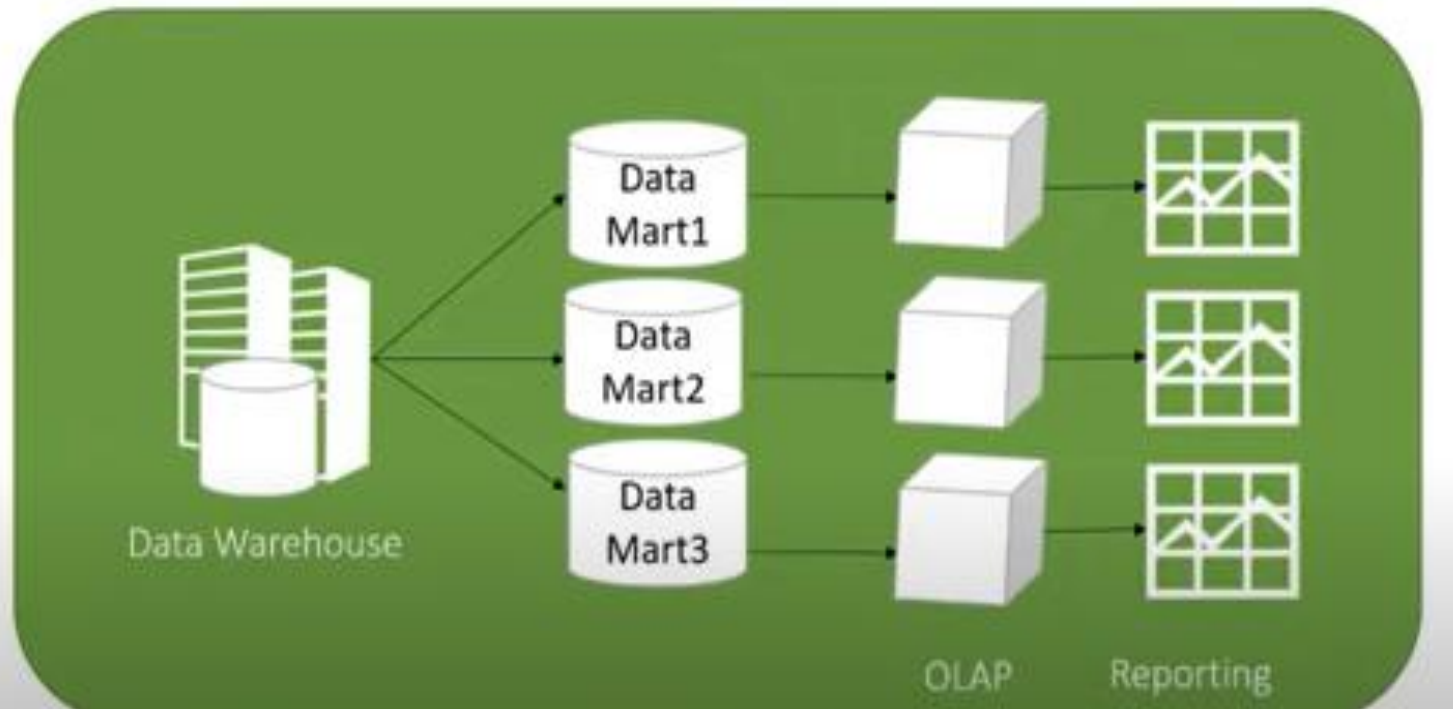
- En el DW los datos actuales son almacenados junto a los datos históricos, y cada dato es marcado con su correspondiente sello de tiempo (timestamps).
- Mediante este sello de tiempo se podrá tener acceso a diferentes versiones de una misma situación.
- Es importante tener en cuenta y planificar la granularidad con que se almacenarán los datos.

 **ene/2016**  **feb/2016**  **mar/2016**  **abr/2016**



ARQUITECTURA DE UN DATA WAREHOUSE: Visión Inmon

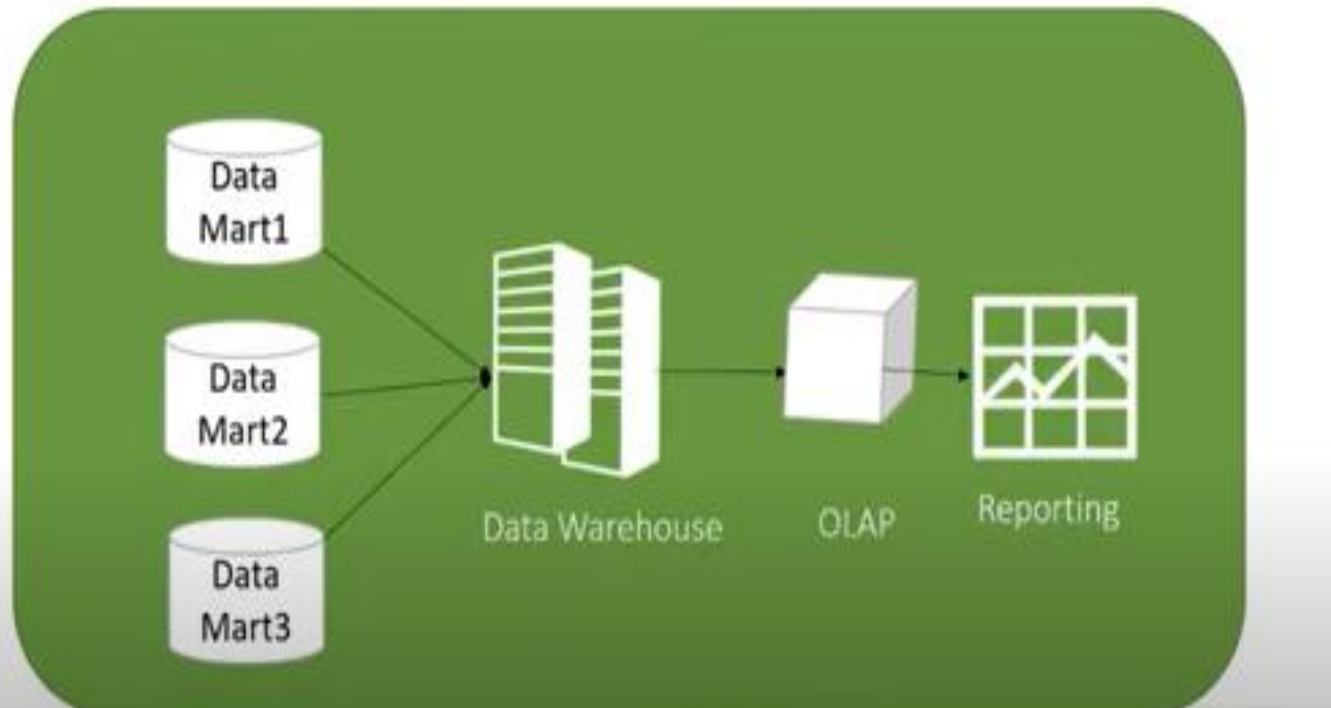
Inmon



En esta visión, el propósito inicial es crear el proyecto de DW, y de allí extraer los Data Marts (Modelo Copo de nieve).

ARQUITECTURA DE UN DATA WAREHOUSE: Visión Kimball

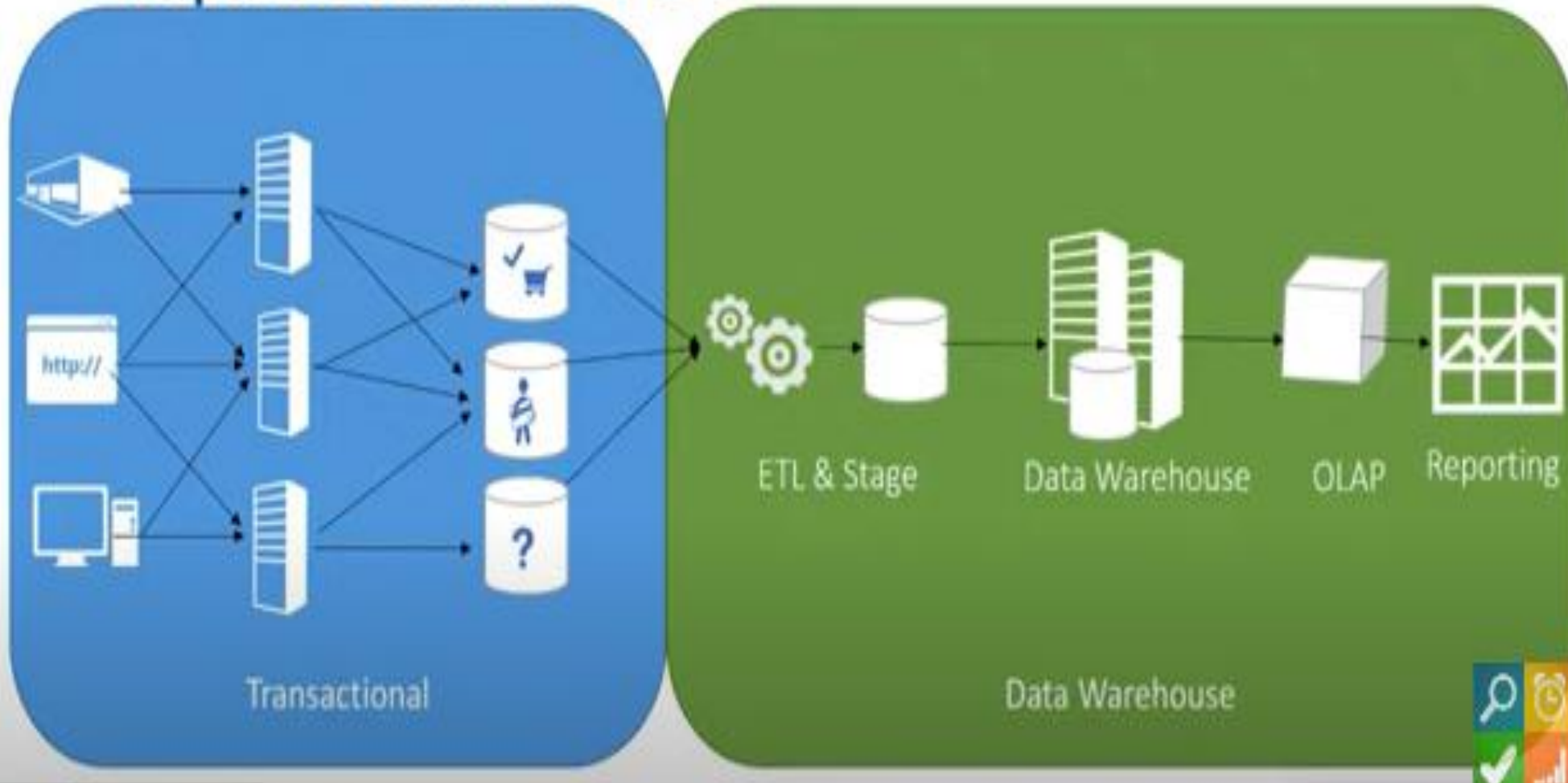
Kimball



El DW es una copia de las transacciones de datos en una parte estructurada para consultas y análisis. En esta visión, el DW es la unión de los Data Marts en una sola entidad. Metodología Ascendente

CAPAS DE LA ARQUITECTURA DE UN DATA WAREHOUSE

Arquitectura del DW



CAPAS DE LA ARQUITECTURA

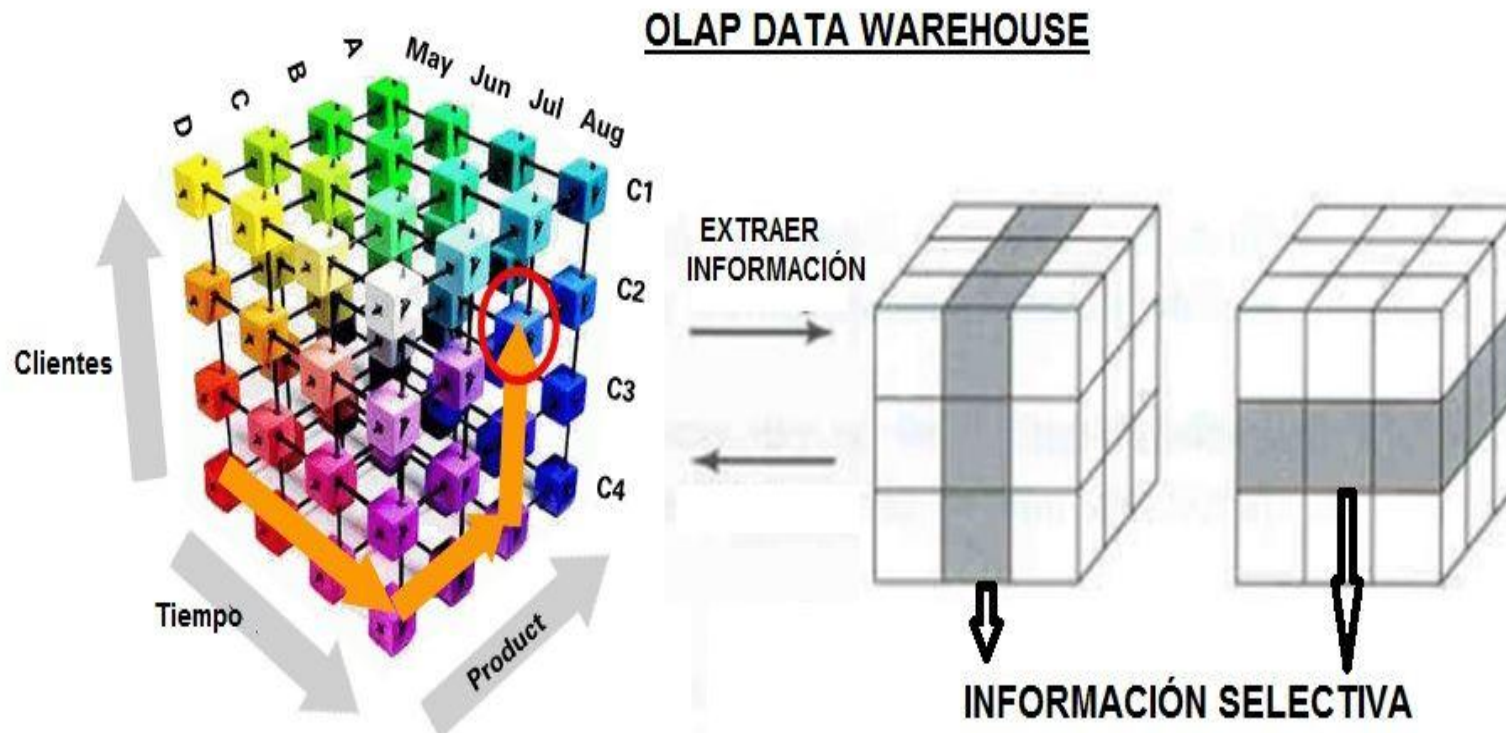
- La capa **Transaccional** está formada por los sistemas transaccionales, web services, OLTP. El objetivo de todos los sistemas se traduce en una base de datos.
- El **Stage** es una copia de lo que originalmente se tiene en los sistemas transaccionales, los que pueden ser de diferentes fuentes (Oracle, SQL Server, Excel, etc.). Tener este almacén temporal permite mayor eficacia de los recursos, y minimizar la posibilidad de pérdida de información ante eventos adversos.
- **ETL (Extract, Transform and Load)** es el motor que hará la carga de los datos, las transformaciones sobre esos datos, la aplicación de las reglas de negocios, la limpieza de los datos, y los moverá finalmente al DW. Herramientas conocidas:
 - Oracle Warehouse Builder
 - IBM Websphere DataStage
 - Microsoft SQL Server Integration Services (SSIS)

CAPAS DE LA ARQUITECTURA

- El **Data Warehouse** es el almacén de datos en sí, junto con las estructuras, métricas y dimensiones definidas. Acá se prevé todo para luego tener la historia de los datos y que sea más sencillo operacionalizarlos.

CAPAS DE LA ARQUITECTURA

- **OLAP (Cubo)**= On-Line Analytical Processing = Procesamiento analítico en línea.
- Es el método más utilizado para analizar y evaluar los datos de la data warehouse en línea. Para analizar los datos se utilizan un conjunto de operaciones.



CAPAS DE LA ARQUITECTURA

- **Data Mining** significa "minería de datos".
- Técnicamente, el datamining es el proceso de encontrar correlaciones entre los patrones o de los campos en grandes bases de datos como en los de una data warehouse.



COMPONENTES A TENER EN CUENTA EN LA IMPLEMENTACIÓN

- **Hardware:** pocos usuarios con unas necesidades muy grandes de información. Sistemas escalables, arquitectura abierta.
- **Software de almacenamiento (SGBD):** relacional o multidimensional
- **Software de extracción y manipulación de los datos**
- **Herramientas de BI**

BENEFICIOS 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; obteniendo un valor añadido para el negocio de dicha información.
- Proporciona la capacidad de **aprender de los datos del pasado y de predecir situaciones futuras** en diversos escenarios.
- 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.

APLICACIÓN EN SECTOR SALUD

- Gestión de la información sanitaria
- Problemática actual
- Soluciones proyectadas
- Debilidades

SOLUCIONES PROYECTADAS: ESTRATEGIA DE INFOESTRUCTURA

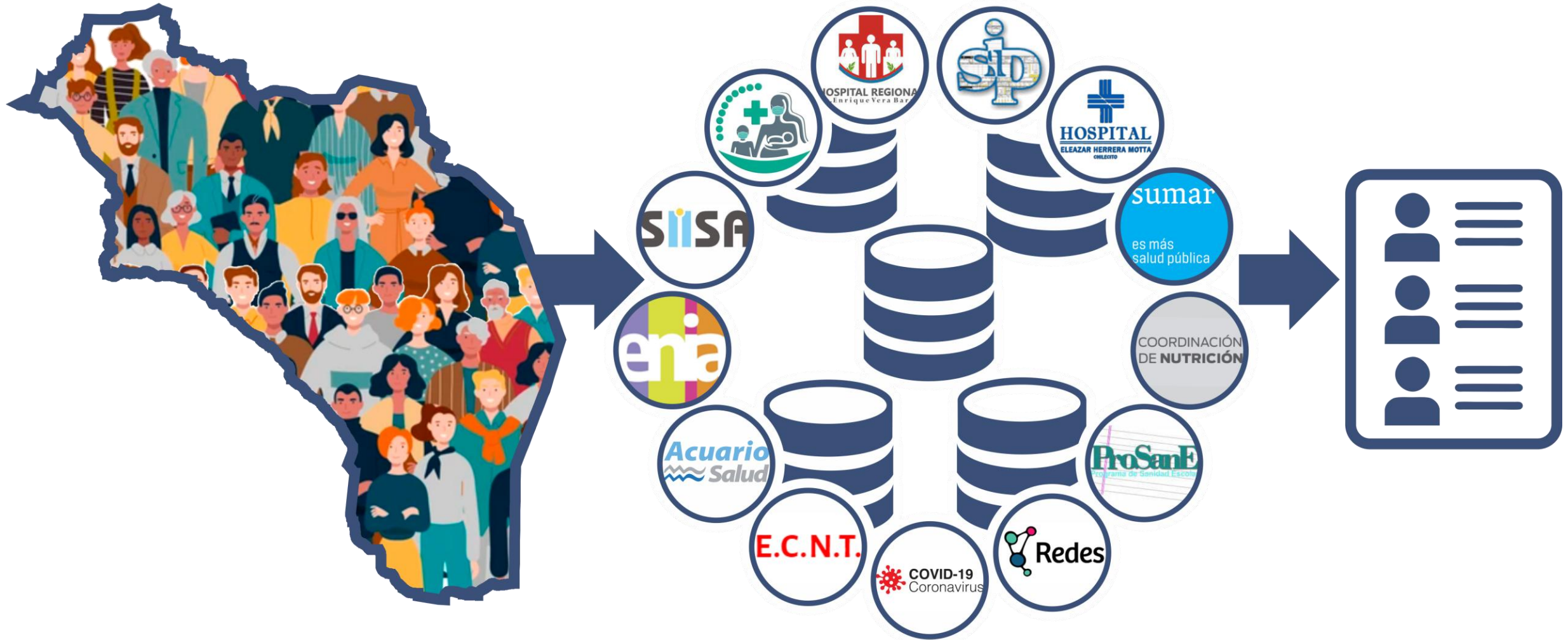
- 1. Padrón único sanitario**
- 2. Set de reportes prestacionales priorizados**

1. PADRÓN ÚNICO SANITARIO

- Empadronamiento de todas las personas que residen en la provincia, a partir de bases de datos existentes (más de 10, ej: Padrón SUMAR, Vacunación, Comunidad Sana, Prestacionales, etc.)
- Variables Priorizadas:
 1. Identificatorias
 2. Domicilio y Georreferenciación
 3. Asignación unívoca a Efector, según domicilio
 4. Personas Gestantes
 5. EAR/RNAR
 6. Cronicidad
 7. Malnutrición
 8. Población cautiva (en Residencias, y privados de libertad)
 9. Cobertura Pública Exclusiva



2. SET DE REPORTES PRESTACIONALES PRIORIZADOS



Ejemplo 1: Reporte 23. Salud Mental

- **ESTRUCTURA:**

- IDENTIFICACIÓN USUARIO
- EFECTOR
- FECHA/TIPO PRESTACIÓN/DIAGNÓSTICO

- **LÍNEAS PRIORIZADAS:**

- INTENTO DE SUICIDIO
- CONSUMO PROBLEMÁTICO DE ALCOHOL
- CONSUMO DE SUSTANCIAS PSICOACTIVAS
- OTRAS ADICCIONES

Ejemplo 1: Reporte 23. Salud Mental

- **FUENTES:**

1. FACTURACIÓN SUMAR

2. ACUARIO

3. HSI

4. CONSULTAS HMYN

5. CONSULTAS HEVB

6. EMERGENCIAS
HOSPITALARIAS

7. BD EGRESOS HOSPITALARIOS

8. BD HOSPITAL DE DIA

9. SAN AGUSTÍN

10. DAS

11. DISMI

12. SERVICIO DE SALUD MENTAL

Ejemplo 1: Reporte 23. Salud Mental

- **AVANCES:**

1. CASI 21.750 PRESTACIONES

2. 6389 PERSONAS, IDENTIFICADOS Y ASIGNADAS A UN EFECTOR. ENTRE ELLAS

- a. INTENTO DE SUICIDIO: 402

- b. CONSUMO PROBLEMÁTICO DE ALCOHOL: 465

- c. CONSUMO DE SUSTANCIAS: 778

DOS NECESIDADES DIFERENTES

Operar

Registrar una consulta.

Registrar una vacuna.

Asignar un turno.

Actualizar un domicilio.

Registrar una internación.

La prioridad es que la transacción se procese correctamente.

OLTP

Analizar

Comparar períodos.

Observar tendencias.

Agrupar por territorio.

Comparar establecimientos.

Analizar diferentes dimensiones.

La prioridad es obtener información para decidir.

OLAP

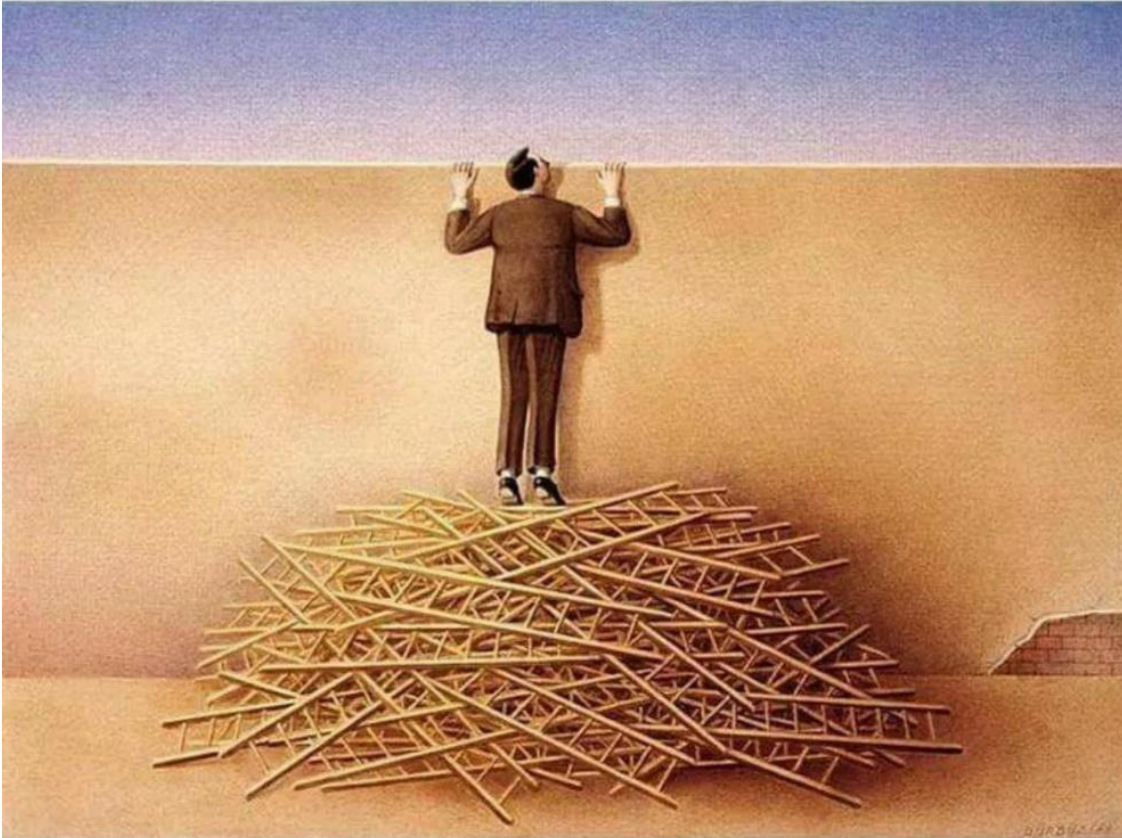
DUDAS? PREGUNTAS? REFLEXIONES?



CONCLUSIONES

- Un Data Warehouse es un sistema, no un producto, en el que se almacenan datos. Es una técnica para consolidar y administrar datos de variadas fuentes con el propósito de responder preguntas de negocios y tomar decisiones, de una forma rápida.
- El Data Warehouse separa la carga del análisis y normalmente contiene datos históricos derivados de datos transaccionales.
- Necesidades no cubiertas de datawarehousing en decisores de políticas de estado. Salud, como ejemplo descripto.

**NO IMPORTA CUÁNTOS
RECURSOS TENGAS**



**SI NO SABES CÓMO UTILIZARLOS
NUNCA SERÁ SUFICIENTE**

¡MUCHAS GRACIAS!