



ESCUELA DE  
INGENIERÍA EN CIENCIAS Y SISTEMAS  
FACULTAD DE INGENIERÍA  
UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



# Evolución de la Arquitectura de Business Intelligence

Del Data Warehouse Tradicional a los Sistemas Analíticos Modernos

IMPARTIDO POR: Piter Valiente de León



## 31 de Marzo

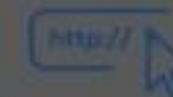
15:40 - 17:20 HORAS



Google Meet



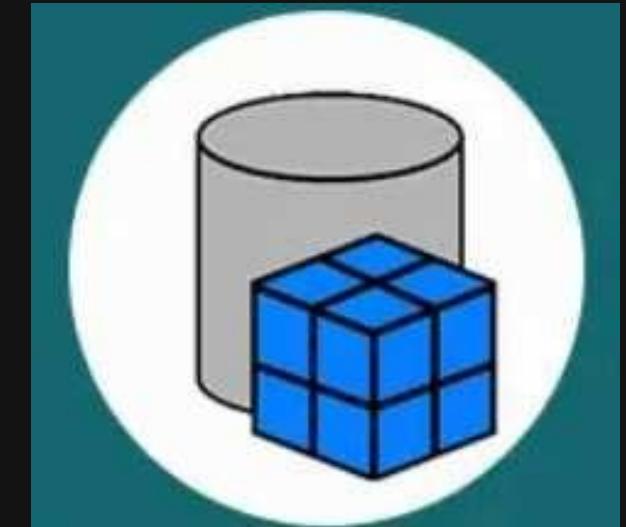
/ecysFIUSAC



<https://dtt-ecys.org>

# Bases de datos multidimensionales

## ¿Qué es?

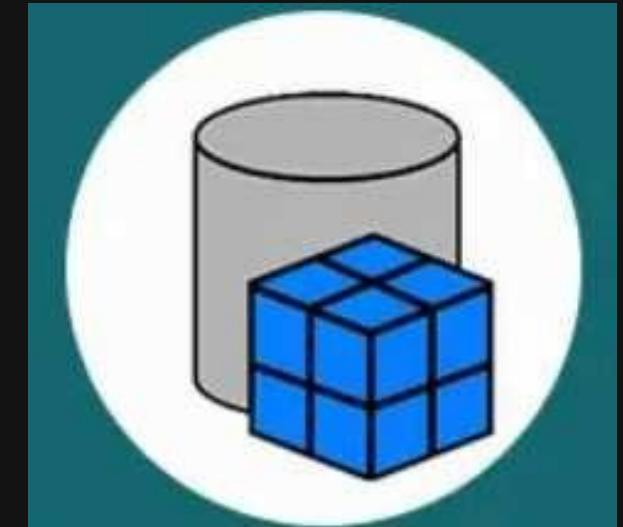


Las bases de datos multidimensionales (MDBs) son un tipo de bases de datos optimizadas para el análisis de datos y el procesamiento de consultas complejas, particularmente en aplicaciones de inteligencia empresarial y análisis de grandes volúmenes de datos.

Se basan en un modelo multidimensional que permite organizar los datos en estructuras llamadas cubos, donde cada dimensión del cubo representa una perspectiva diferente desde la cual los datos pueden ser visualizados y analizados.

# Bases de datos multidimensionales

## Modelo de datos multidimensional



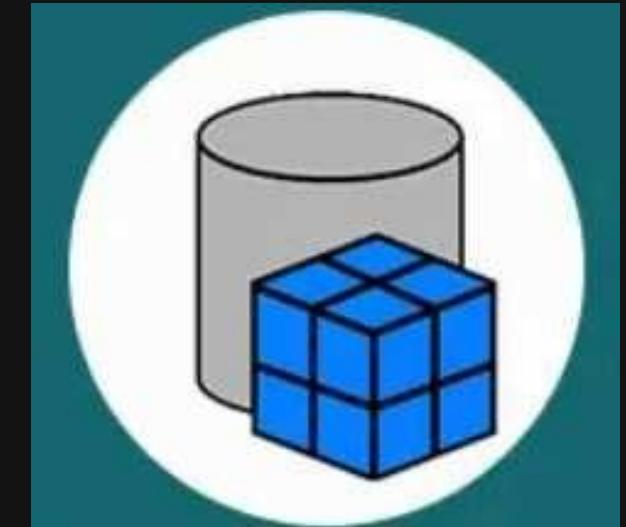
Se basan en un modelo multidimensional que permite organizar los datos en estructuras llamadas cubos, donde cada dimensión del cubo representa una perspectiva diferente desde la cual los datos pueden ser visualizados y analizados.

Los datos se organizan en cubos que contienen múltiples dimensiones. Estas dimensiones pueden ser cualquier atributo de los datos, como el tiempo, la ubicación geográfica, el producto, o el cliente.

Las dimensiones permiten a los usuarios visualizar y analizar los datos desde diferentes ángulos o niveles de agregación.

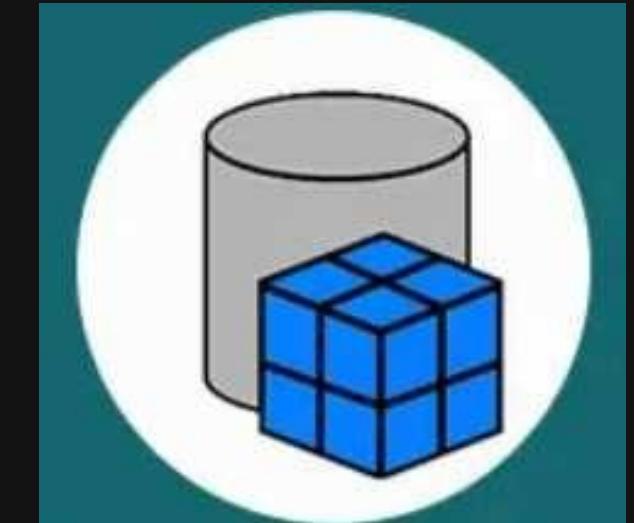
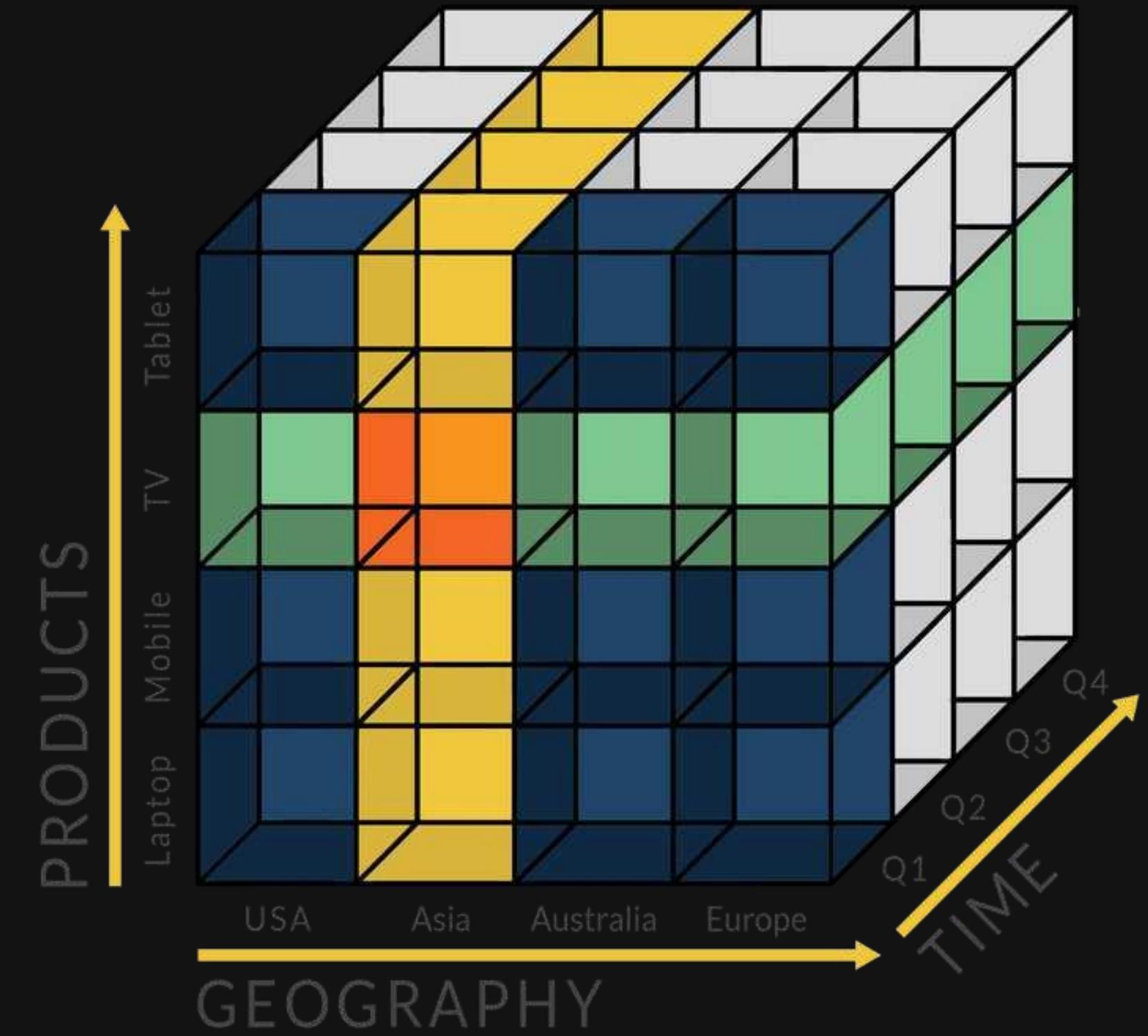
# Bases de datos multidimensionales

## Cubo de datos



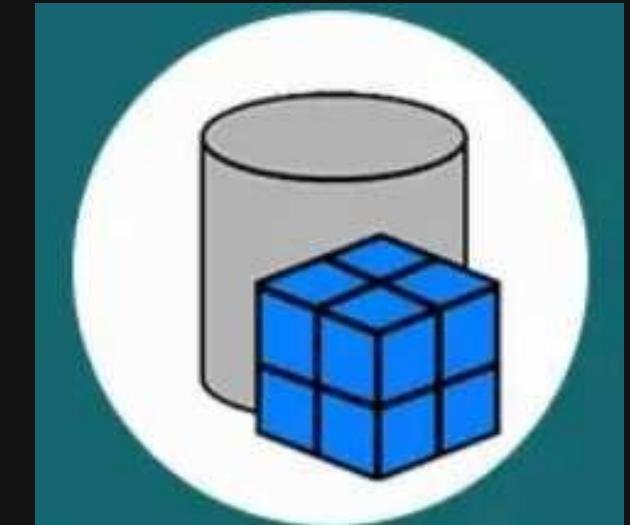
Los datos en una MDB se almacenan en un formato cúbico, con cada celda del cubo representando una combinación única de dimensiones y conteniendo medidas o métricas cuantitativas (como ventas, ingresos, costos, etc.).

Los cubos permiten realizar operaciones rápidas de drill-down (descomposición de los datos en niveles más detallados) y roll-up (resumen de los datos a niveles más altos de agregación).



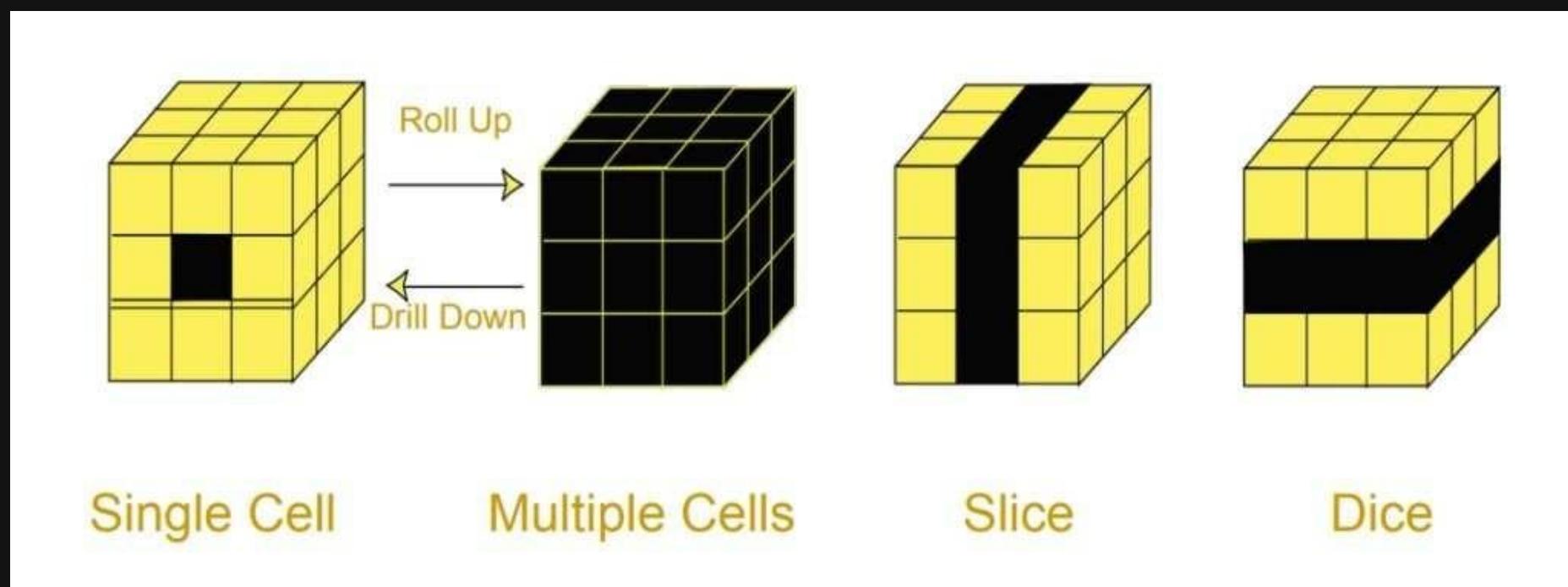
# Bases de datos multidimensionales

## Procesamiento analítico en línea (OLAP)



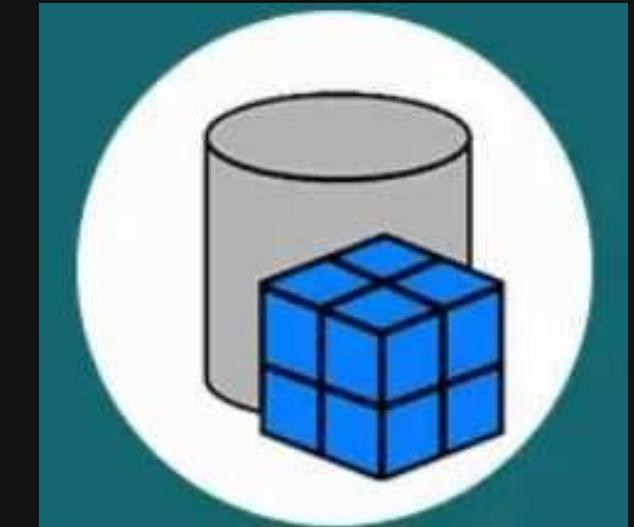
Las bases de datos multidimensionales son la base del procesamiento analítico en línea (Online Analytical Processing u OLAP), que permite realizar consultas ad hoc de manera eficiente y analizar grandes volúmenes de datos rápidamente.

OLAP utiliza operaciones como slice and dice (filtrar y reorganizar datos), drill-down y roll-up.



# Bases de datos multidimensionales

## Expresiones Multidimensionales (MDX)



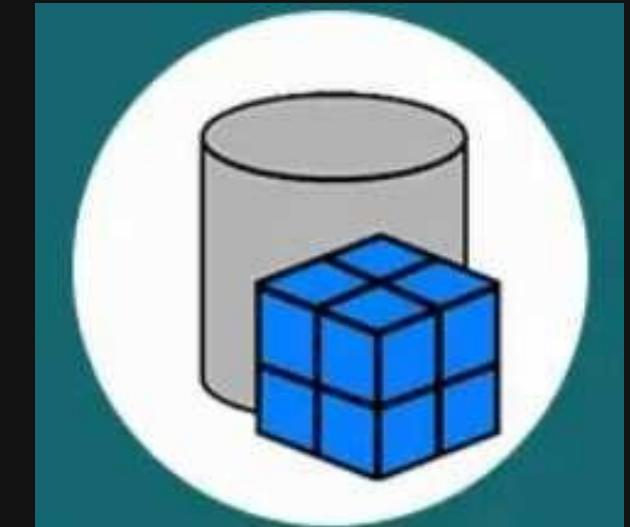
MDX (Multidimensional Expressions) es un lenguaje de consulta diseñado para interactuar con bases de datos multidimensionales, especialmente en entornos OLAP.

MDX es un estándar de facto utilizado principalmente para realizar consultas sobre cubos OLAP y obtener información agregada desde diferentes dimensiones.

```
SELECT
    [Dimensión Tiempo].[Año].Members ON COLUMNS, -- Sección de columnas
    [Dimensión Producto].[Producto].Members ON ROWS -- Sección de filas
FROM
    [Cubo Ventas] -- Cubo del cual se extraen los datos
WHERE
    ([Medida Ventas]) -- Medida que se quiere analizar
```

# Bases de datos multidimensionales

## Rendimiento optimizado

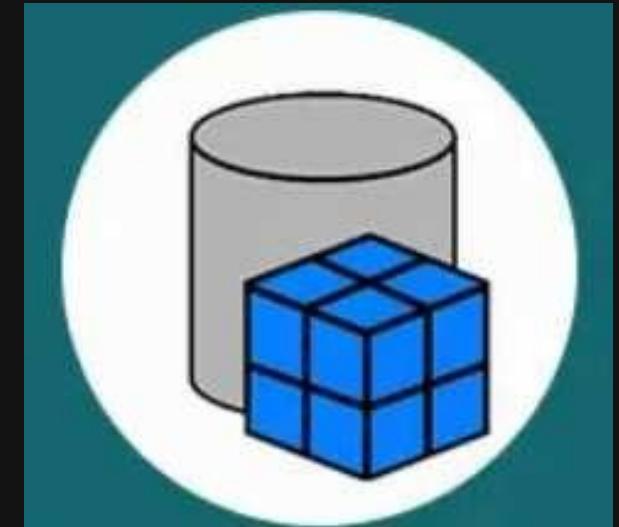


Las MDBs están optimizadas para consultas de lectura, lo que las hace ideales para el análisis de datos y reportes, ya que permiten realizar análisis en tiempo real sobre grandes volúmenes de datos de manera eficiente.

La estructura multidimensional reduce la complejidad de las consultas en comparación con las bases de datos relacionales tradicionales.

# Bases de datos multidimensionales

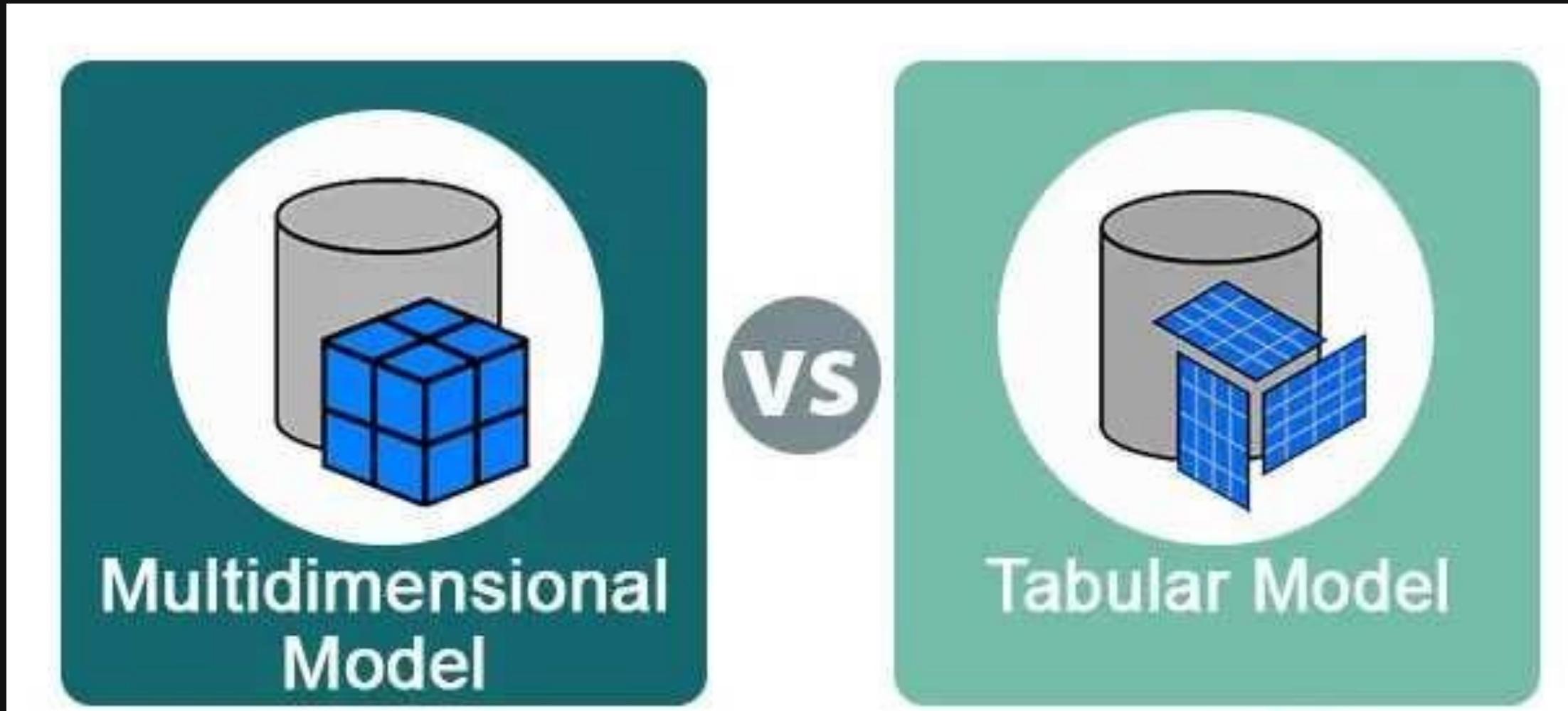
## Usos comunes



- Análisis financiero y contable: Permite realizar análisis de ingresos, gastos y presupuestos desde múltiples perspectivas (por ejemplo, por departamento, por mes, por producto).
- Inteligencia empresarial (BI): Facilita la creación de informes y paneles de control para la toma de decisiones.
- Análisis de ventas y marketing: Permite descomponer y analizar datos de ventas según diversas dimensiones como región, producto y tiempo.

# Bases de datos multidimensionales

## Modelo relacional vrs Modelo multidimensional

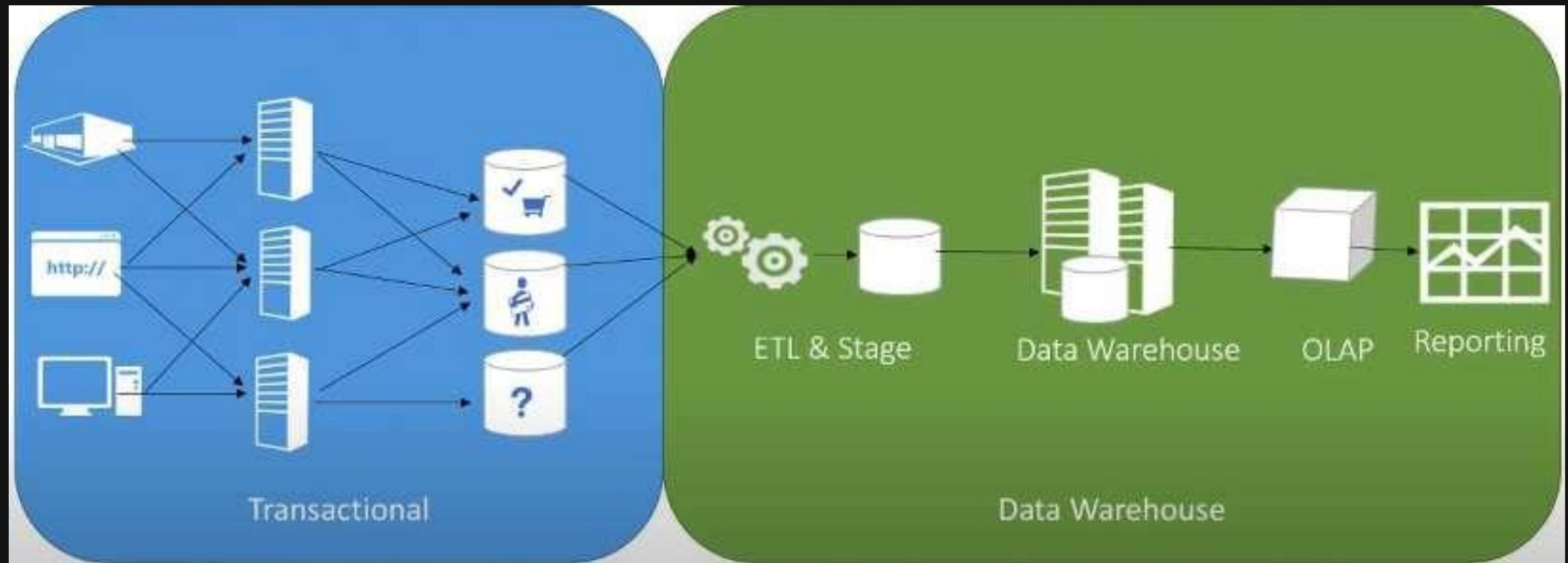
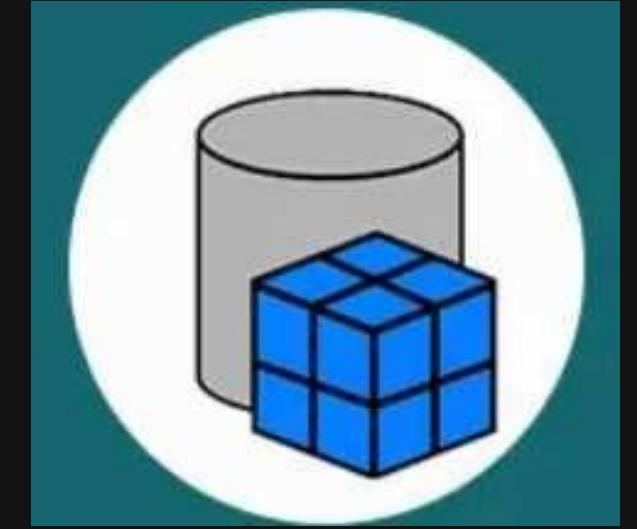


# Bases de datos multidimensionales

**Modelo relacional vrs Modelo multidimensional**


# Bases de datos multidimensionales

## Fases para crear un cubo



# Extracción, Transformación y Carga (ETL)

## ¿Qué es?

El proceso ETL (Extract, Transform, Load) es un conjunto de pasos fundamentales en la integración de datos, especialmente en sistemas de data warehouses.

ETL permite recopilar datos de diferentes fuentes, transformarlos según las necesidades del negocio, y cargarlos en un repositorio central, como un almacén de datos (data warehouse), para su análisis y uso.

# ETL

## Extracción

La fase de extracción se encarga de tomar los datos desde diversas fuentes, que pueden ser bases de datos relacionales, archivos planos, sistemas ERP, hojas de cálculo, APIs, etc. La extracción puede implicar la recopilación de datos en diferentes formatos y estructuras.

# ETL

## Transformación

En esta fase, los datos extraídos se transforman para adecuarlos a los requisitos del sistema de destino. Las transformaciones pueden incluir limpieza, formateo, agregación, conversión de tipos de datos, y cálculos específicos para normalizar los datos.

# ETL

## Transformación

### Tipos de Transformaciones:

- **Limpieza de datos:** Eliminación de duplicados, corrección de errores, y llenado de valores faltantes.
- **Conversión de formatos:** Los datos pueden ser transformados a un formato estándar que el sistema de destino entienda.
- **Normalización/Denormalización:** Ajustar la estructura de los datos para hacerlos consistentes y optimizar su uso en análisis.
- **Cálculos y derivaciones:** Generación de nuevos datos o medidas a partir de los datos existentes, como el cálculo de promedios, sumas o tasas.

# ETL

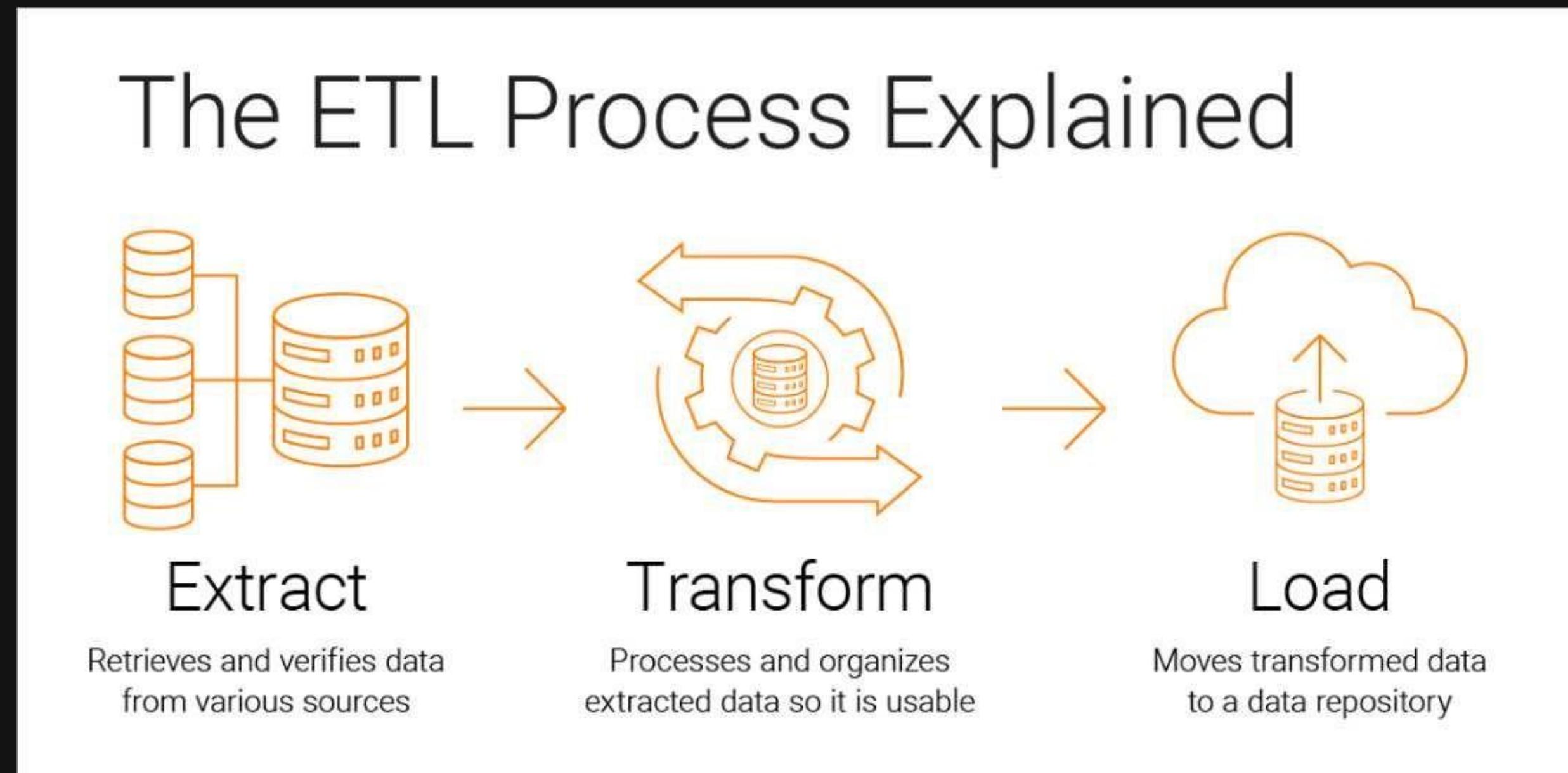
## Carga

### Tipos de Transformaciones:

- **Limpieza de datos:** Eliminación de duplicados, corrección de errores, y llenado de valores faltantes.
- **Conversión de formatos:** Los datos pueden ser transformados a un formato estándar que el sistema de destino entienda.
- **Normalización/Denormalización:** Ajustar la estructura de los datos para hacerlos consistentes y optimizar su uso en análisis.
- **Cálculos y derivaciones:** Generación de nuevos datos o medidas a partir de los datos existentes, como el cálculo de promedios, sumas o tasas.

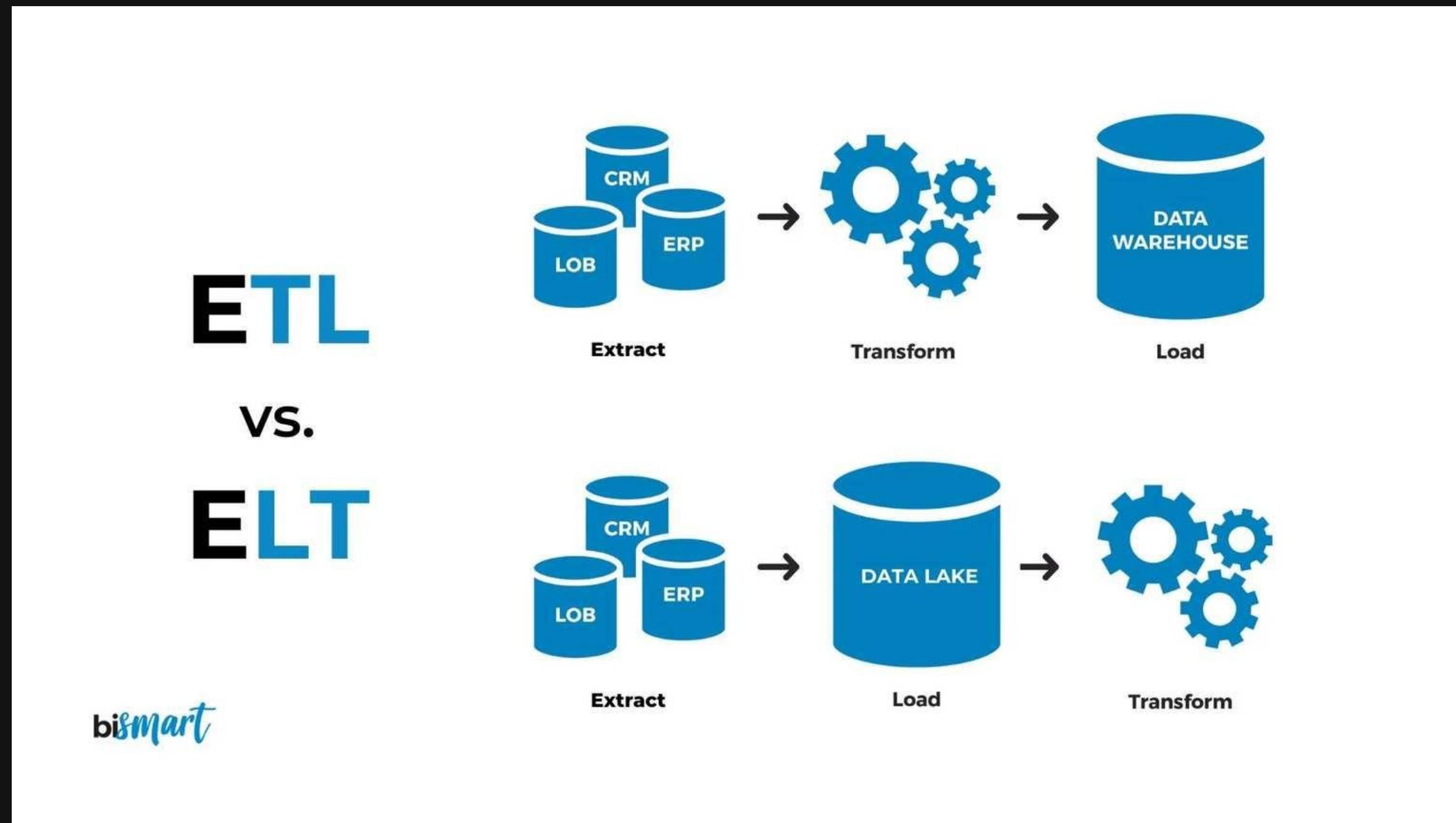
# Extracción, Transformación y Carga (ETL)

## ¿Qué es?



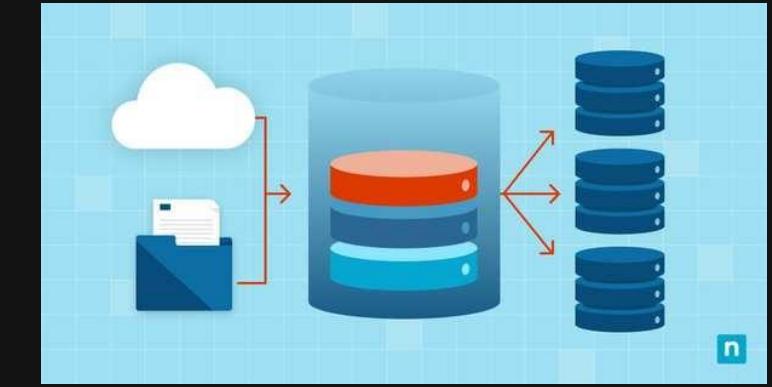
# Extracción, Transformación y Carga (ETL)

¿Qué es?



# Data Warehouse

## ¿Qué es?



Un Data Warehouse (almacén de datos) es un sistema de almacenamiento diseñado específicamente para la consulta y el análisis de grandes volúmenes de datos, provenientes de múltiples fuentes. Su objetivo principal es centralizar y organizar la información en un formato estructurado para facilitar la toma de decisiones en una organización.

# Data Warehouse

## Orientado a temas



Los datos se organizan por temas o áreas de interés, como ventas, finanzas, o clientes, en lugar de estar organizados por transacciones u operaciones cotidianas. Esto permite que los usuarios accedan a la información relevante de manera más eficiente para análisis específicos.

# Data Warehouse

## Datos históricos



A diferencia de las bases de datos operacionales que almacenan datos actuales y transaccionales, un Data Warehouse conserva grandes volúmenes de datos históricos. Estos permiten analizar tendencias y realizar comparaciones a lo largo del tiempo.

# Data Warehouse

## Esquemas de organización de datos



### Esquema estrella:

Centraliza una tabla de hechos (datos transaccionales o métricas) rodeada de varias tablas dimensionales (dimensiones como tiempo, cliente, producto).



# Data Warehouse

## Esquemas de organización de datos

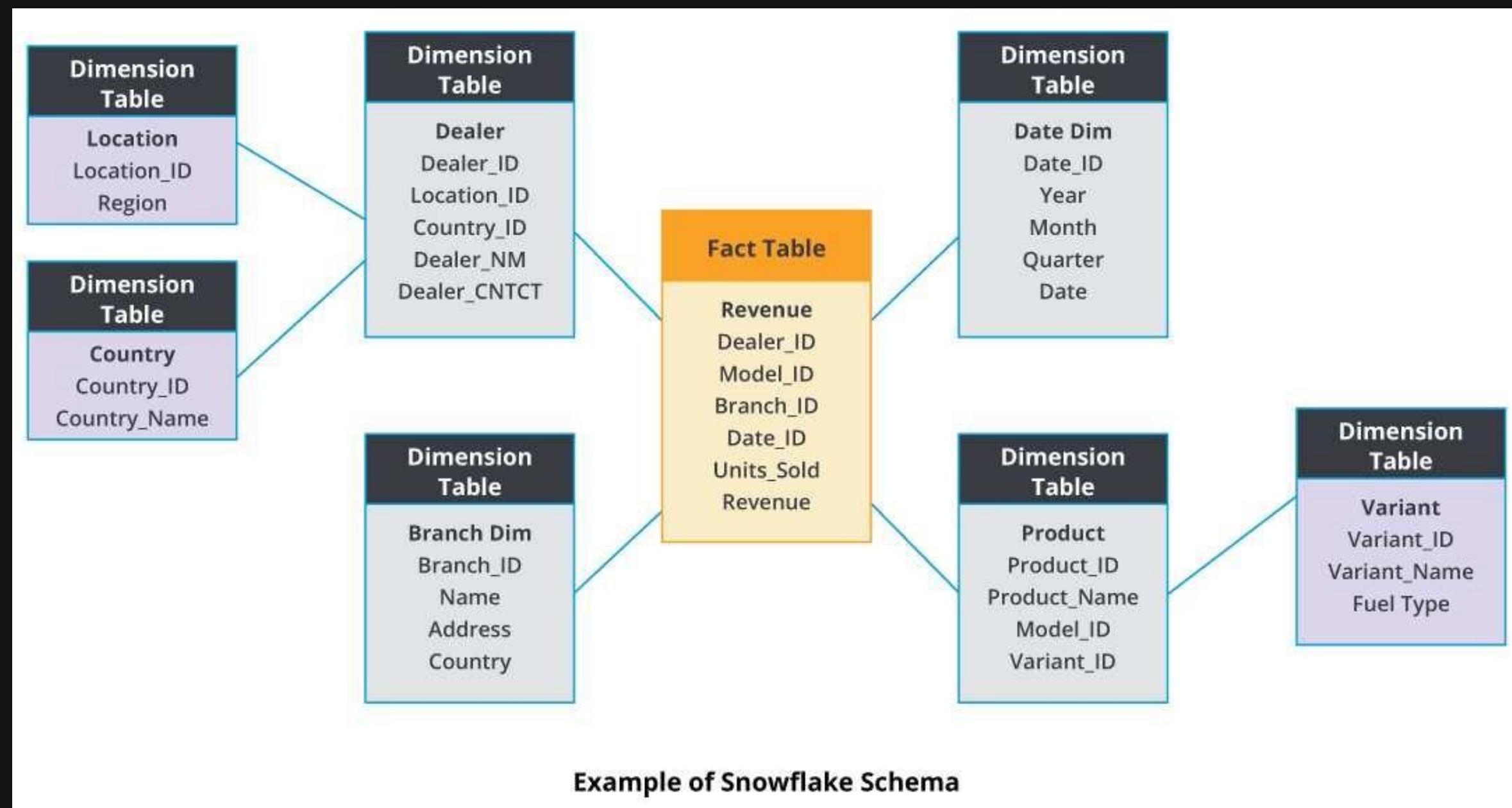


### Esquema copo de nieve:

Es una variante del esquema estrella donde las dimensiones están normalizadas, lo que significa que se dividen en más tablas, reduciendo la redundancia.

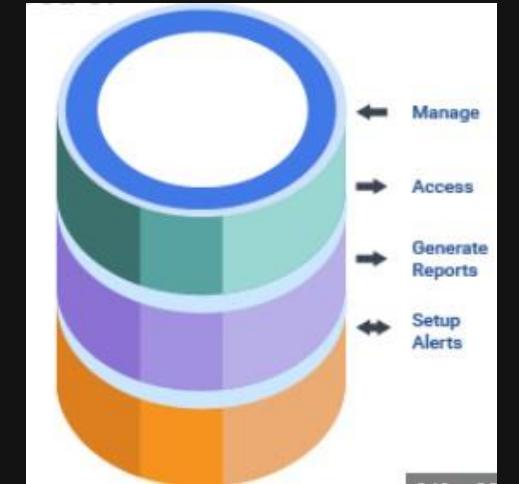
# Data Warehouse

## Esquemas de organización de datos



# Data Mart

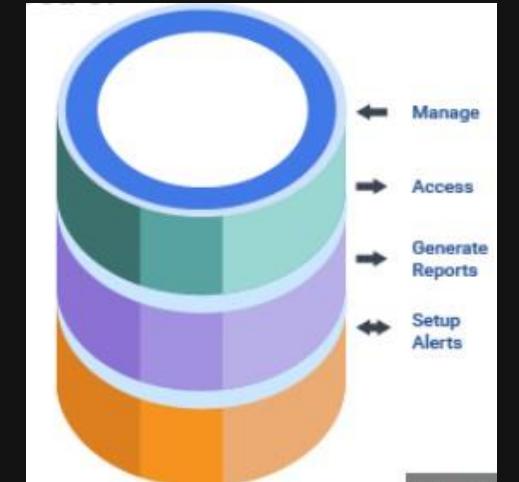
## ¿Qué es?



Un datamart es un subconjunto de un datawarehouse que está diseñado para atender las necesidades específicas de una línea de negocio o de un grupo de usuarios. Almacena datos que son relevantes para un área en particular, como ventas, finanzas o marketing, y está optimizado para el análisis de información dentro de ese contexto.

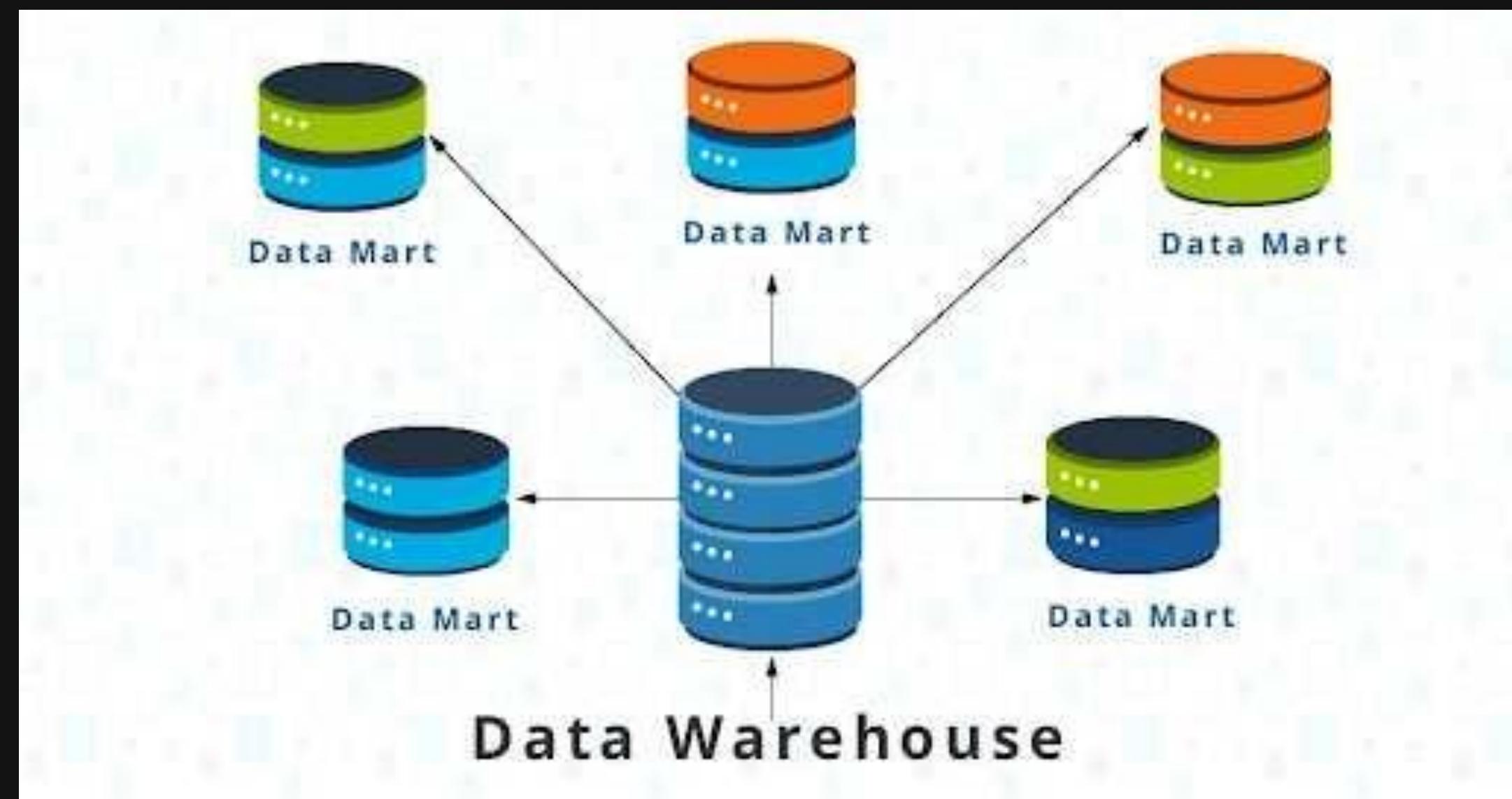
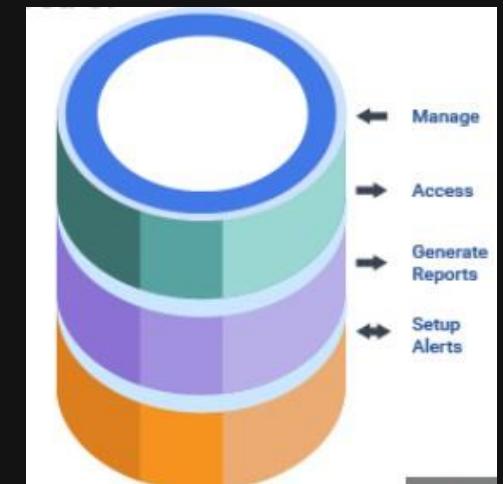
# Data Mart

## Propósito

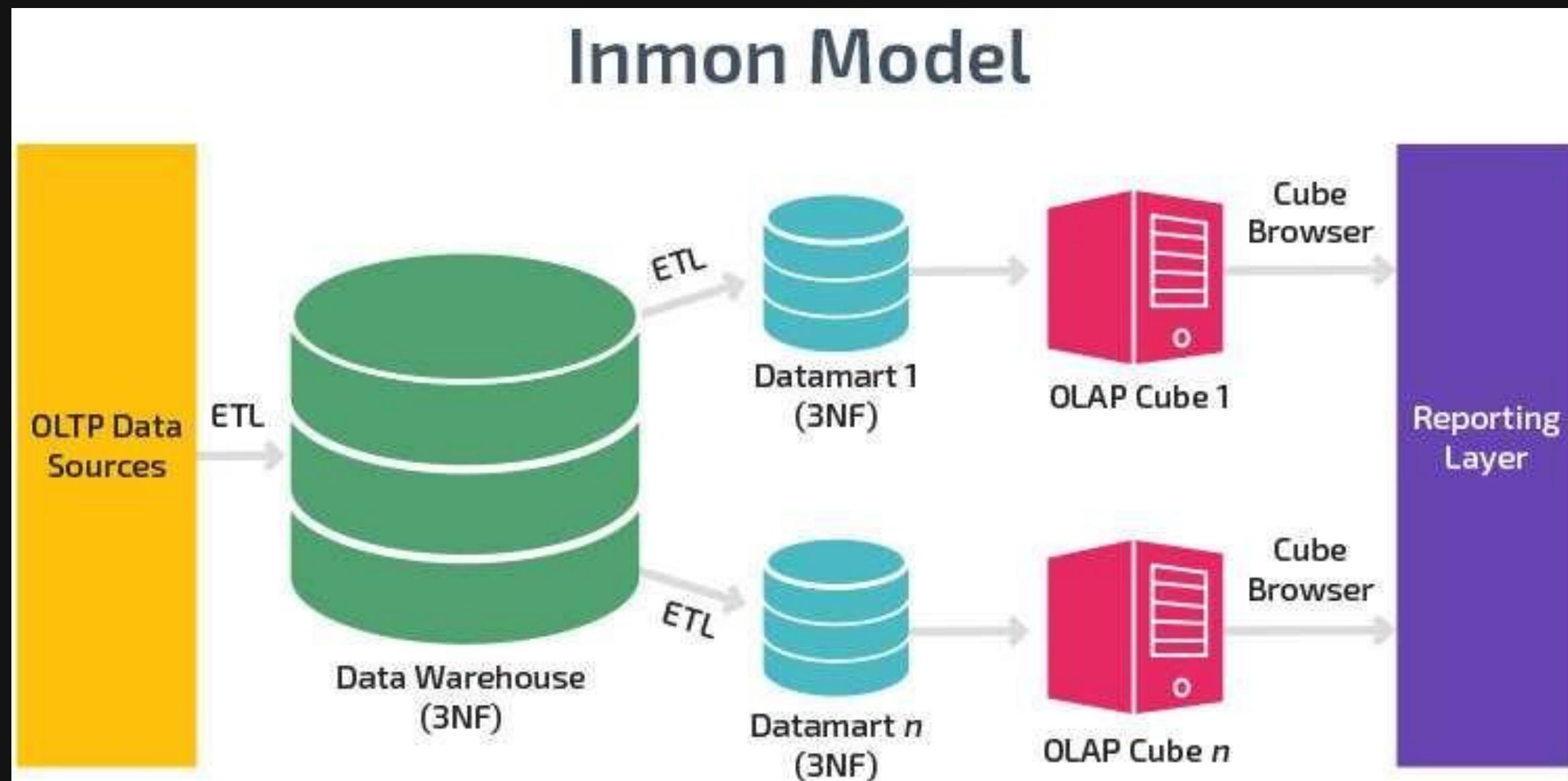
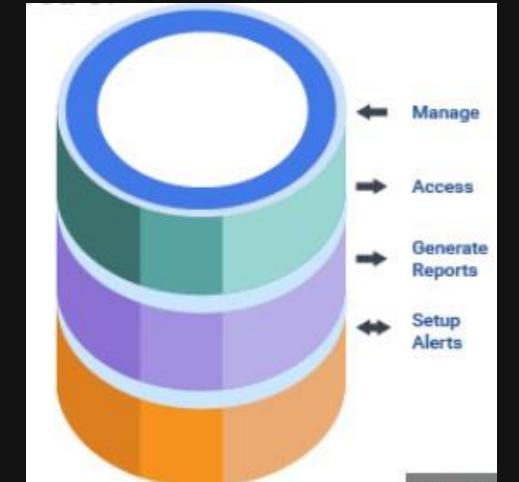


- Facilitar el acceso rápido y eficiente a datos relevantes para departamentos o equipos específicos.
- Ofrecer datos preprocesados y organizados, optimizados para consultas específicas o necesidades de análisis.
- Reducir la complejidad al aislar los datos de interés de una área particular.

# Data Mart



# Data Mart



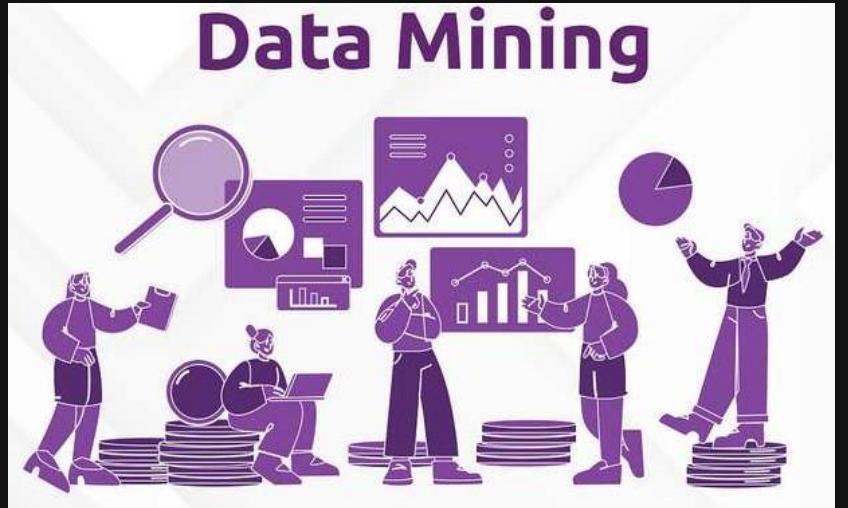
# Pongámoslo a prueba!



# Data Mining

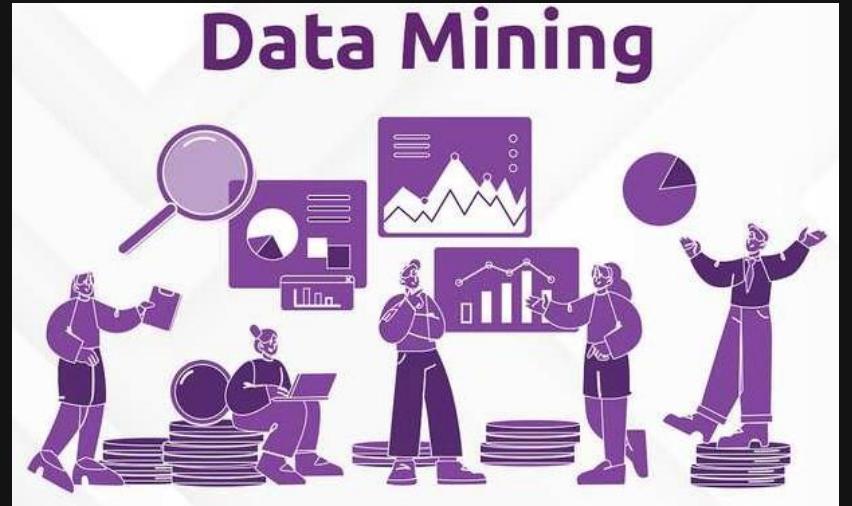
## ¿Qué es?

El Data Mining (minería de datos) es el proceso de analizar grandes conjuntos de datos para descubrir patrones, relaciones y tendencias ocultas que no son evidentes a simple vista.



# Data Mining

## Datos



Los datos son el conjunto de información cruda que se recopila de diversas fuentes. Pueden ser estructurados (como bases de datos), semiestructurados (como archivos XML) o no estructurados (como textos, imágenes o videos).

# Data Mining

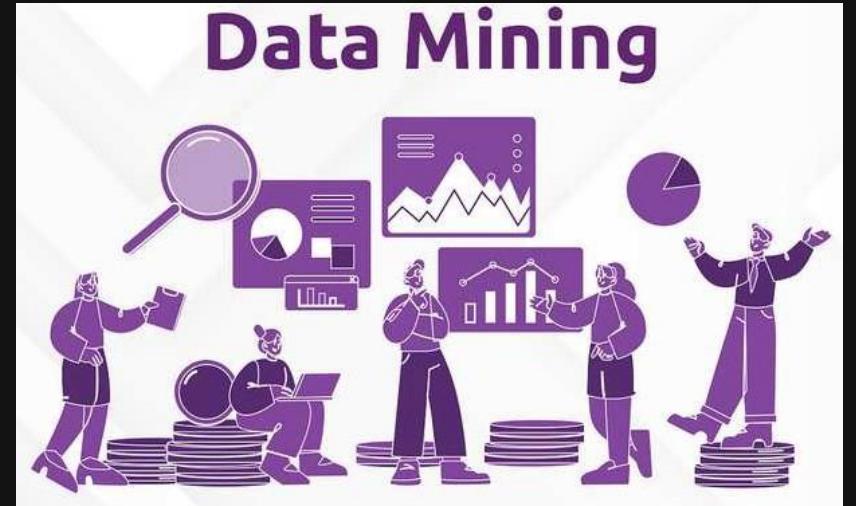
## Minearía de datos



Es el proceso de explorar grandes cantidades de datos para identificar patrones o tendencias relevantes que puedan ser utilizados para la toma de decisiones. Utiliza técnicas de estadística, inteligencia artificial, y machine learning.

# Data Mining

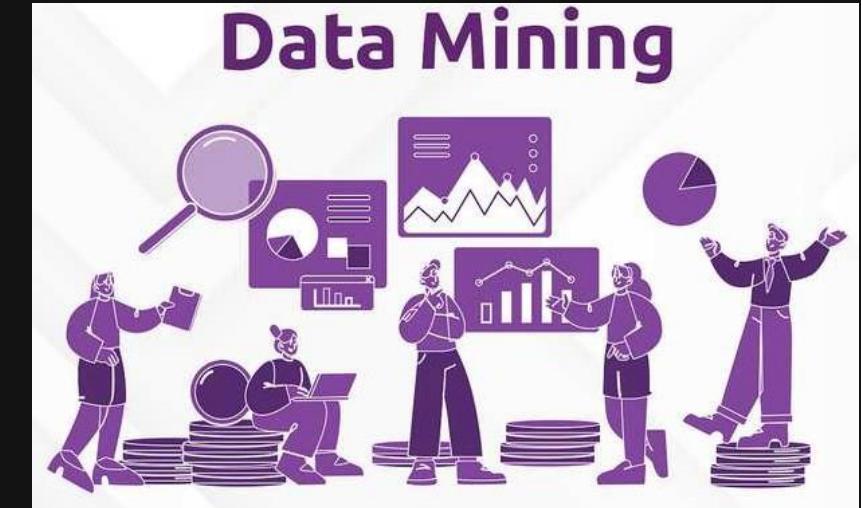
## Dataset



Es la colección de datos sobre la cual se aplica el análisis. Un dataset normalmente está organizado en forma de una tabla, donde las columnas representan atributos (características) y las filas representan las instancias (datos).

# Data Mining

## Modelos de Data Mining

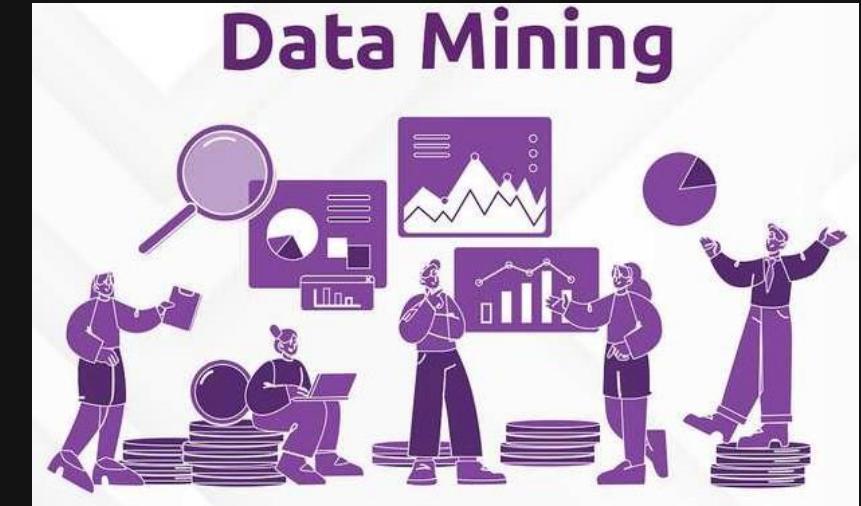


Son los algoritmos o técnicas que se aplican a los datos para identificar patrones. Los principales modelos son:

- Clasificación: Asignar etiquetas a los datos basándose en un conjunto de categorías (Ej.: Spam o No Spam).
- Regresión: Predecir valores continuos (Ej.: Predecir ventas futuras).
- Agrupamiento (Clustering): Agrupar datos en subconjuntos que son más similares entre sí que con los datos de otros grupos (Ej.: Segmentación de clientes).

# Data Mining

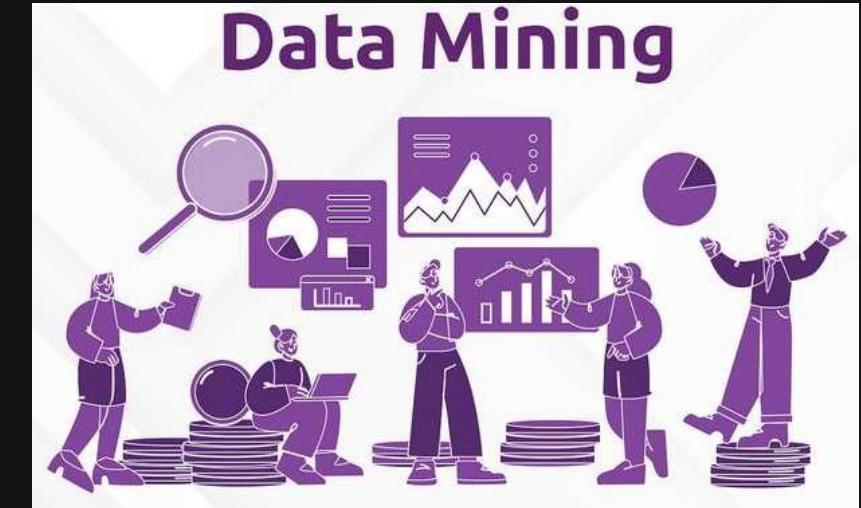
## Modelos de Data Mining



- Reglas de Asociación: Descubrir relaciones o asociaciones entre elementos (Ej.: "Los clientes que compran X también tienden a comprar Y").
- Análisis de Secuencia: Identificar patrones secuenciales o dependencias entre eventos (Ej.: Modelar el comportamiento de un usuario en una web).

# Data Mining

## Algoritmos de Data Mining



- K-means (Clustering): Agrupa datos en 'k' clusters o grupos.
- Árboles de decisión (Clasificación): Crea un modelo basado en reglas que divide los datos en categorías.
- Regresión lineal (Regresión): Modelo estadístico para predecir el valor de una variable en función de otras.
- Apriori (Reglas de Asociación): Encuentra relaciones frecuentes entre elementos en bases de datos transaccionales.

# ¿Dudas?

