

# Introducción a **DATA MINING** (Minería de Datos)



Tecnología Superior en Big Data

**MINERÍA DE DATOS**

Ciclo: M3A

# Introducción a Data Mining

## DEFINICION

La minería de datos es el proceso de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente, estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiado datos.

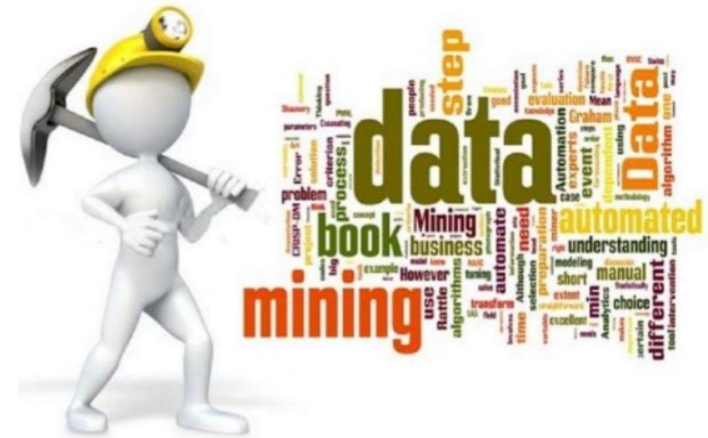
Estos patrones y tendencias se pueden recopilar y definir como un modelo de minería de datos. Los modelos de minería de datos se pueden aplicar en escenarios como los siguientes:

- |  |  |
|--|--|
| <b>Escenarios para<br/>Modelos de Minería<br/>de Datos</b> | <ul style="list-style-type: none"><li>• <b>Pronostico</b></li><li>• <b>Calculo de Riesgo y Probabilidad</b></li><li>• <b>Recomendaciones</b></li><li>• <b>Búsqueda de secuencias</b></li><li>• <b>Agrupación</b></li></ul> |
|--|--|

# Introducción a Data Mining

## OBJETIVO

El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior. Además de la etapa de análisis en bruto, supone aspectos de gestión de datos y de bases de datos, de procesamiento de datos, del modelo y de las consideraciones de inferencia, de métricas de Intereses, de consideraciones de la Teoría de la complejidad computacional, de post-procesamiento de las estructuras descubiertas, de la visualización y de la actualización en línea minería de datos es una forma innovadora de obtener información comercial



# Introducción a Data Mining

## ¿QUE NO ES MINERIA DE DATOS?

La Minería de Datos no es agrupar cantidades masivas de información.

El uso de herramientas para ordenar metadatos no elimina la necesidad de conocer la organización y la información de la misma.. Para usar Minería de datos se debe conocer la organización o entidad que quiere la información y que genera los datos.

Los patrones encontrados por la MD deben de ser verificados en el mundo real.

Aunque es una buena herramienta la MD requiere de un buen análisis estadístico y requiere conocimientos sobre el tema en el que se busca información.

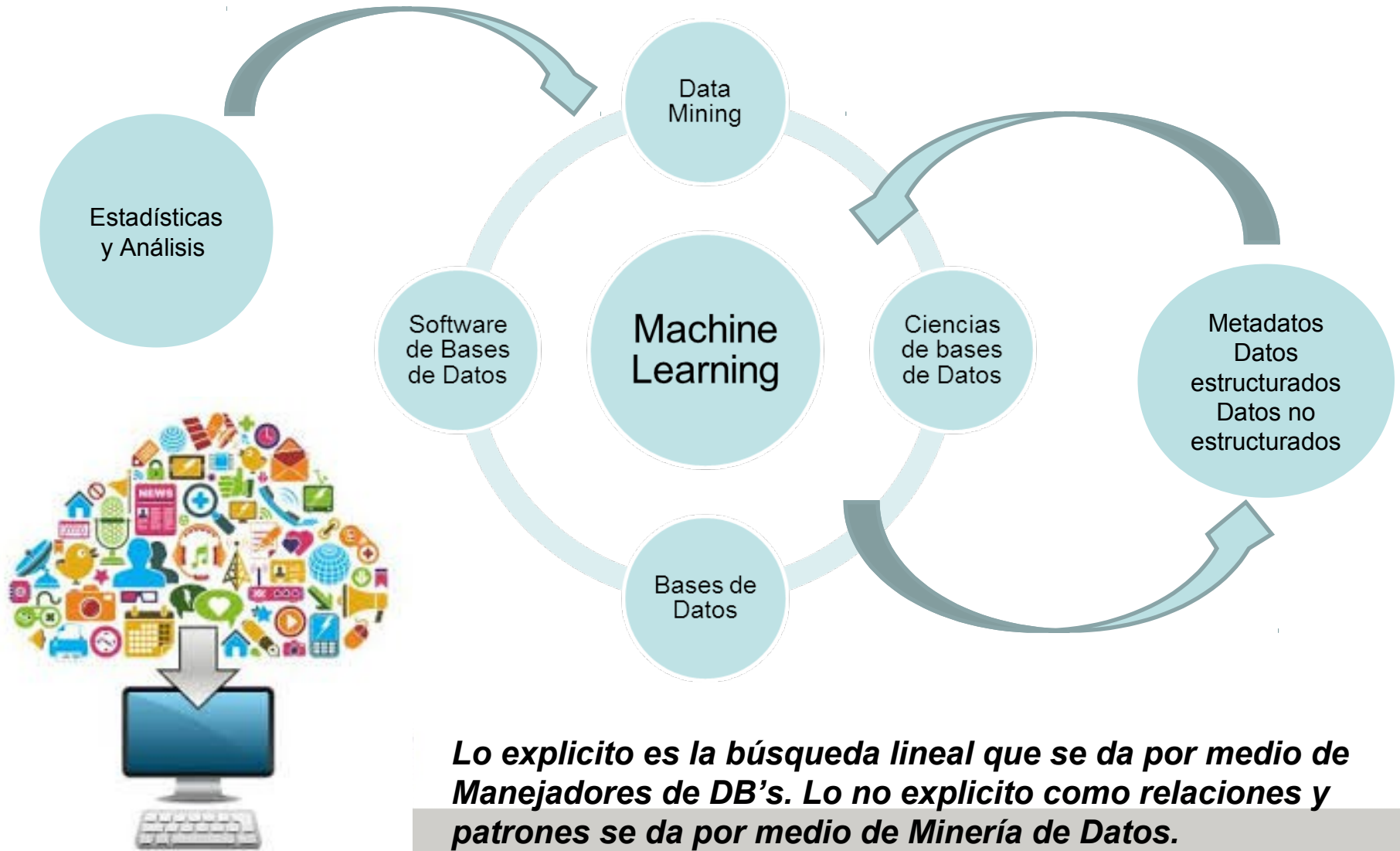
# Introducción a Data Mining

## CIENCIAS COMPUTACIONALES



- **Arquitectura de Computadoras**
- **Bases de Datos**
- **Computo Científico**
- **Redes de Computadoras y TICs**
- **Sistemas Distribuidos, paralelos y concurrentes**
- **Ingeniería de Software**
- **Inteligencia Artificial**
  - **Robótica**
  - **Machine Learning**
  - **Imagenología**
  - **Redes neuronales**
  - **Aprendizaje cognitivo**
  - **Data Mining**

# Introducción a Data Mining



# Introducción a Data Mining

La minería de datos ayuda a...

- Asistir al ANALISIS DE NEGOCIOS
- Ayuda en la toma de decisiones y a mejorar su nivel y fiabilidad de respuesta
- Encontrar Patrones y Relaciones en los datos que pueden ser útiles a una organización sobre algún tópico o sobre el universo de datos.
- Encontrar RELACIONES PREDICTIVAS sobre comportamientos.

Por otro lado la minería de datos ...:

- Debe descartar los datos NO RELEVANTES en el ANALISIS DE DATOS.
- Las RELACIONES PREDICTIVAS no necesariamente son causas de una acción o comportamiento.
- Los ALGORITMOS son sensibles a la selección o exclusión de los datos.
- La Minería de Datos no descubre soluciones automáticamente sin una guía o conocimiento de la entidad que se analiza

# Introducción a Data Mining

## EJEMPLOS DE USO DE MINERIA DE DATOS

Los métodos convencionales pueden indicar a un banco cuál es el tipo de cuenta bancaria más rentable de entre las que proporciona. En cambio, la minería de datos permite al banco crear perfiles de los clientes que ya disponen de ese tipo de cuenta. El banco puede luego utilizar la minería de datos para encontrar otros clientes que coinciden con ese perfil, y así poder emprender una campaña comercial dirigida específicamente a esos clientes.

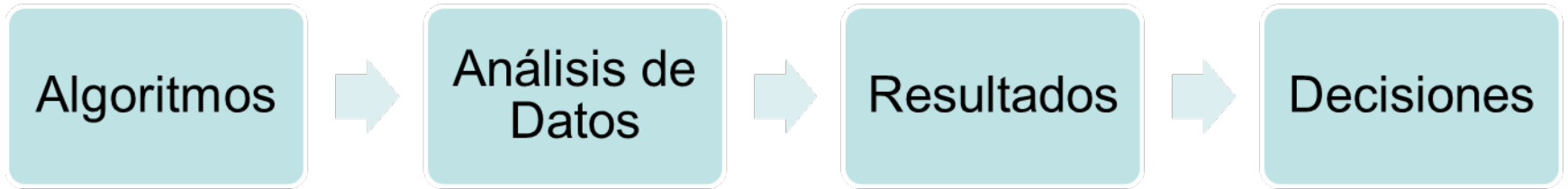
La minería de datos puede identificar patrones en los datos de la empresa, por ejemplo, en los registros de compra de un supermercado. Si, por ejemplo, los clientes compran los productos A y B, ¿qué producto C es más probable que compren también?



***¡Responder con precisión a preguntas como éstas son una ayuda muy valiosa para crear estrategias comerciales!***



# Introducción a Data Mining



***Se requiere un conocimiento de la entidad y de los datos que se buscan.***

***La Minería de Datos no da soluciones, muestra información que ya esta ahí y ayuda a la toma de decisiones.***

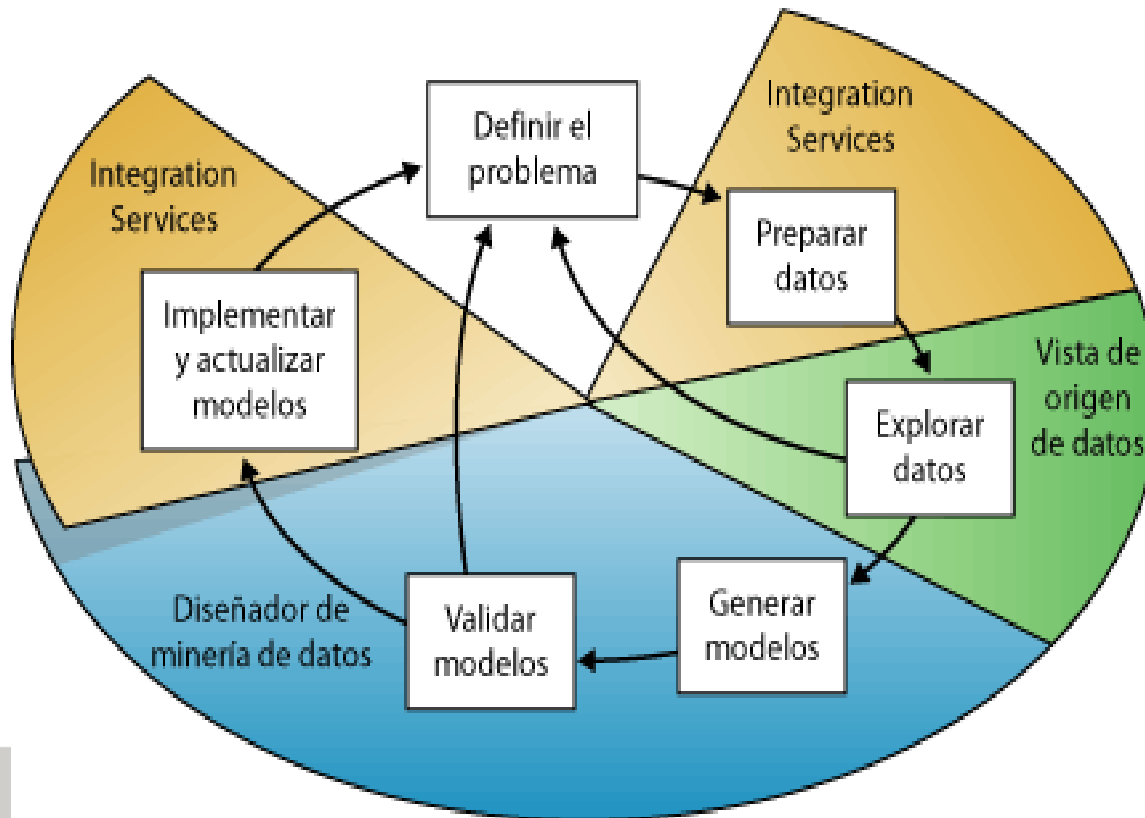
***Es vital que se entienda lo que se busca en los datos. Si los datos no son concisos entonces es que faltan mas datos o el algoritmo esta mal diseñado.***



# Introducción a Data Mining

## PROCESO DE MINERIA DE DATOS

El siguiente diagrama describe las relaciones existentes entre cada paso del proceso y las tecnologías disponibles como las de Microsoft SQL Server que se pueden usar para completar cada paso.

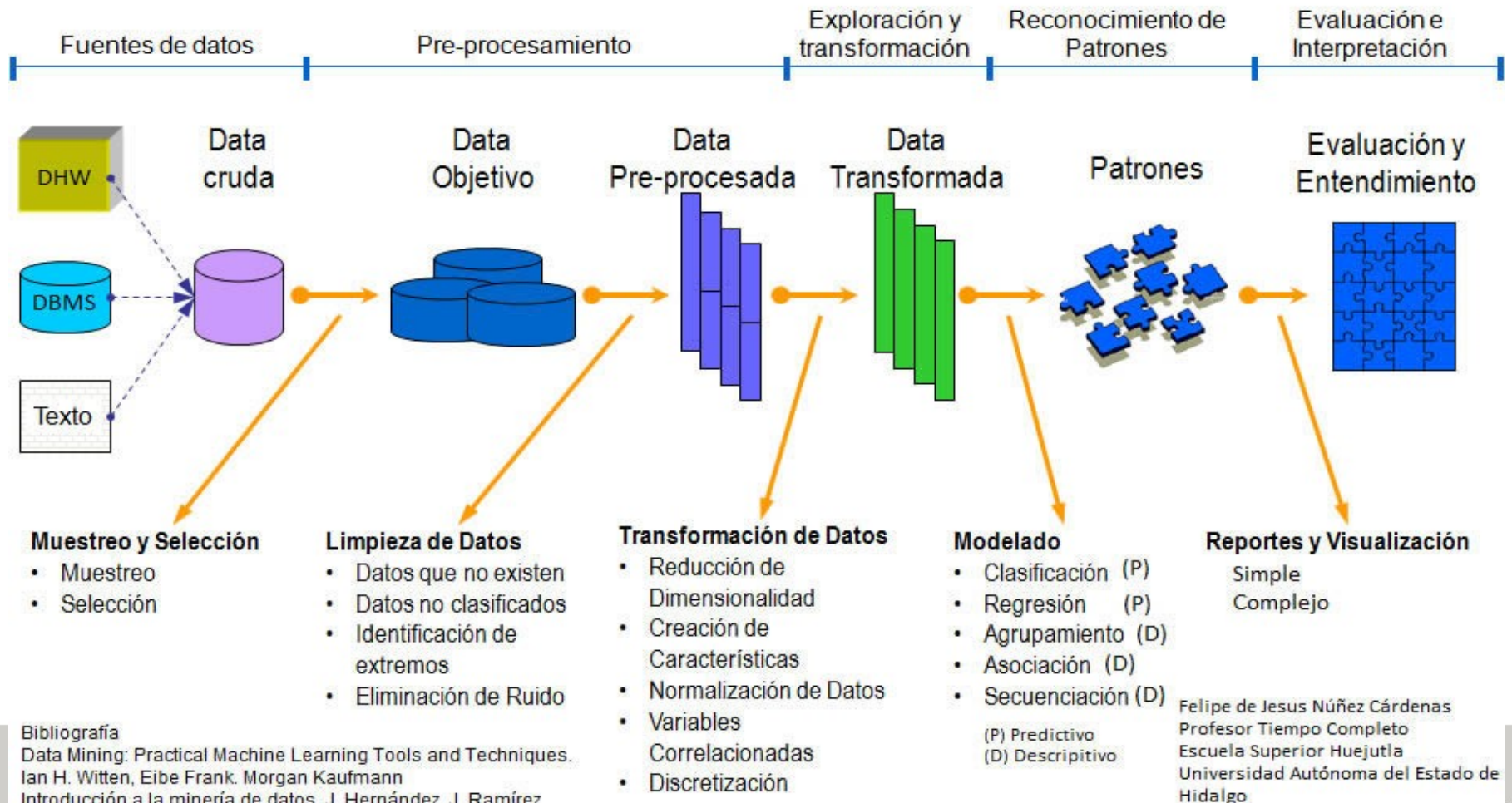


### Pasos para el Proceso de Minería de Datos

1. Definir el problema
2. Preparar datos
3. Explorar los datos
4. Generar un modelo
5. Validar el modelo
6. Implementar y actualizar el modelo

# Introducción a Data Mining

## *El Proceso de la Minería de Datos*



# Introducción a Data Mining

## TÉCNICAS DE MINERÍA DE DATOS

Como ya se ha comentado, las técnicas de la minería de datos provienen de la inteligencia artificial y de la estadística, dichas técnicas, no son más que algoritmos, más o menos sofisticados que se aplican sobre un conjunto de datos para obtener unos resultados.

Las técnicas más representativas son:

- **Redes neuronales.-** Son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida. Algunos ejemplos de red neuronal son:
  - **El perceptrón.**
  - **El perceptrón multicapa.**
- Los mapas auto organizados, también conocidos como **redes de Kohonen**

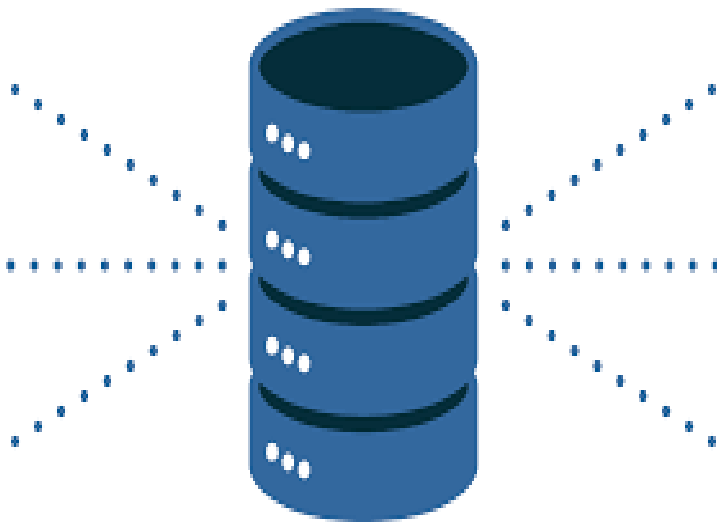
# Introducción a Data Mining

## DATAWAREHOUSE

Viene del Inglés **DATA** = **datos** y **WAREHOUSE** = **almacén**, por eso, una **DATA WAREHOUSE** es un almacén de datos.

Esta palabra se utiliza para hablar de un almacén de datos diseñado para permitir las actividades de inteligencia de un negocio. En definitiva sirve para ayudar analizar los datos recopilados por la empresa con el fin de mejorar su rendimiento.

Un Data Warehouse o Almacén de Datos, es una gran base de datos, normalmente medida en gigabytes (miles de millones de caracteres) o terabytes (billones de letras), que recoge información de múltiples fuentes, y que su actividad se centra en la Toma de Decisiones, es decir, en el análisis de la información, en vez de su captura.



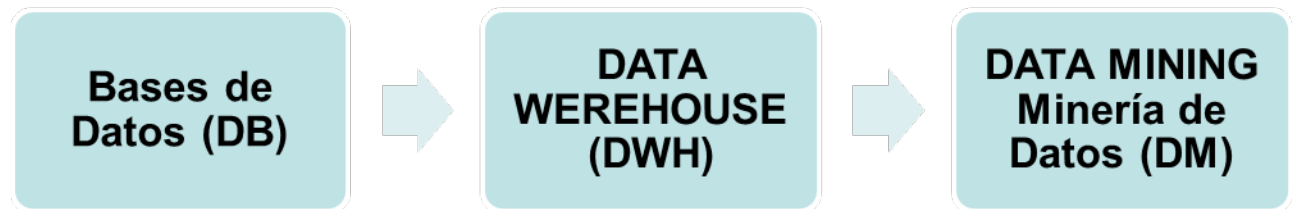
# Introducción a Data Mining

## DATAWAREHOUSE

El Datawarehouse es una colección de datos orientada a un tópico integrado, no volátil y organizado en base al tiempo, o sea con contenido histórico y referenciado.

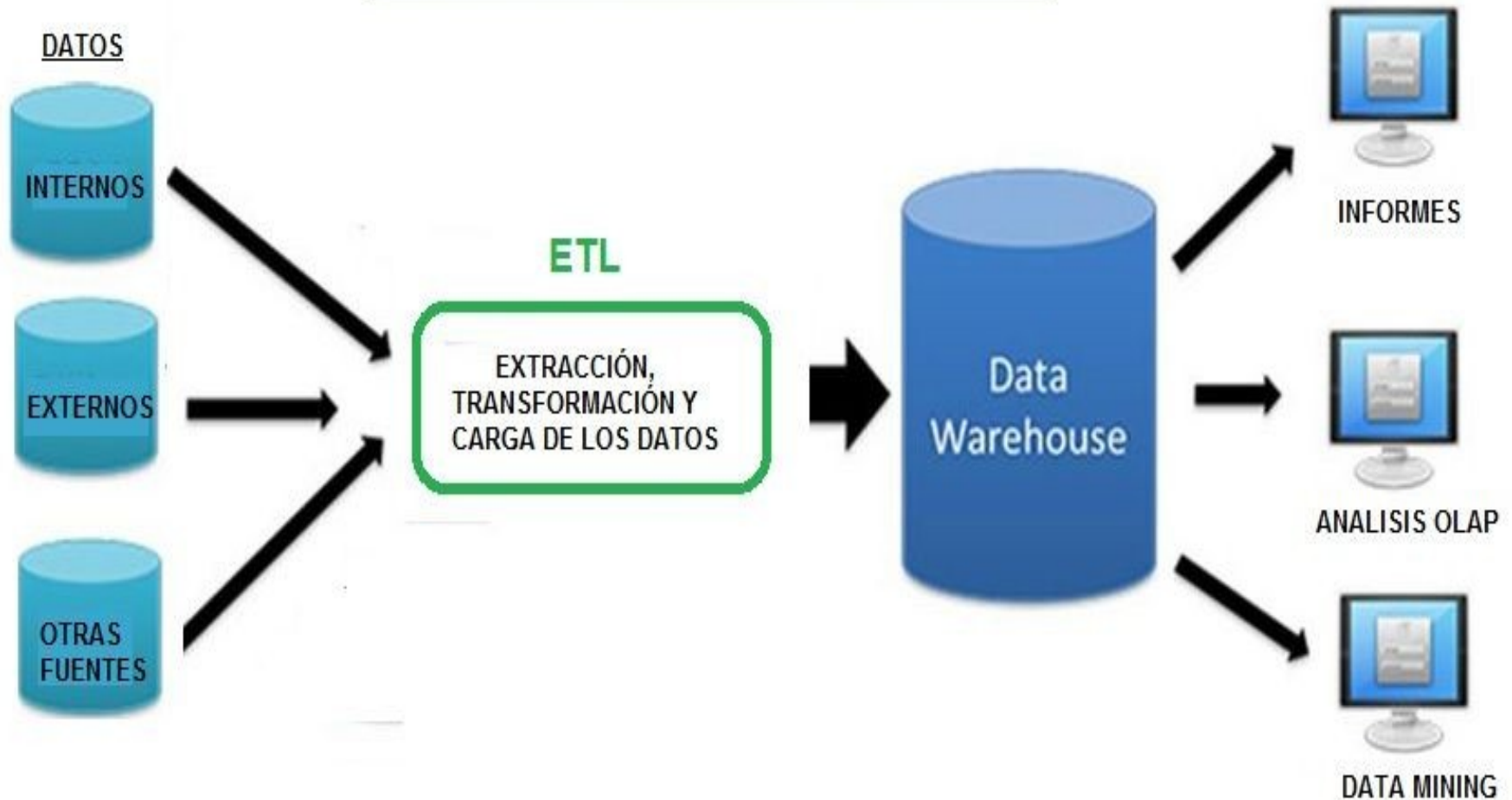
Las características de la información de un Datawarehouse son:

- **INTEGRADA:** Integra datos de diferentes fuentes pero con una sola forma de identificarlo.
- **ATEMPORAL:** Con respaldo y referencia contra el tiempo.
- **NO VOLATIL:** No cambian con el tiempo, debe de ser concisas y fiables.



# Introducción a Data Mining

## DATA WAREHOUSE ARQUITECTURA

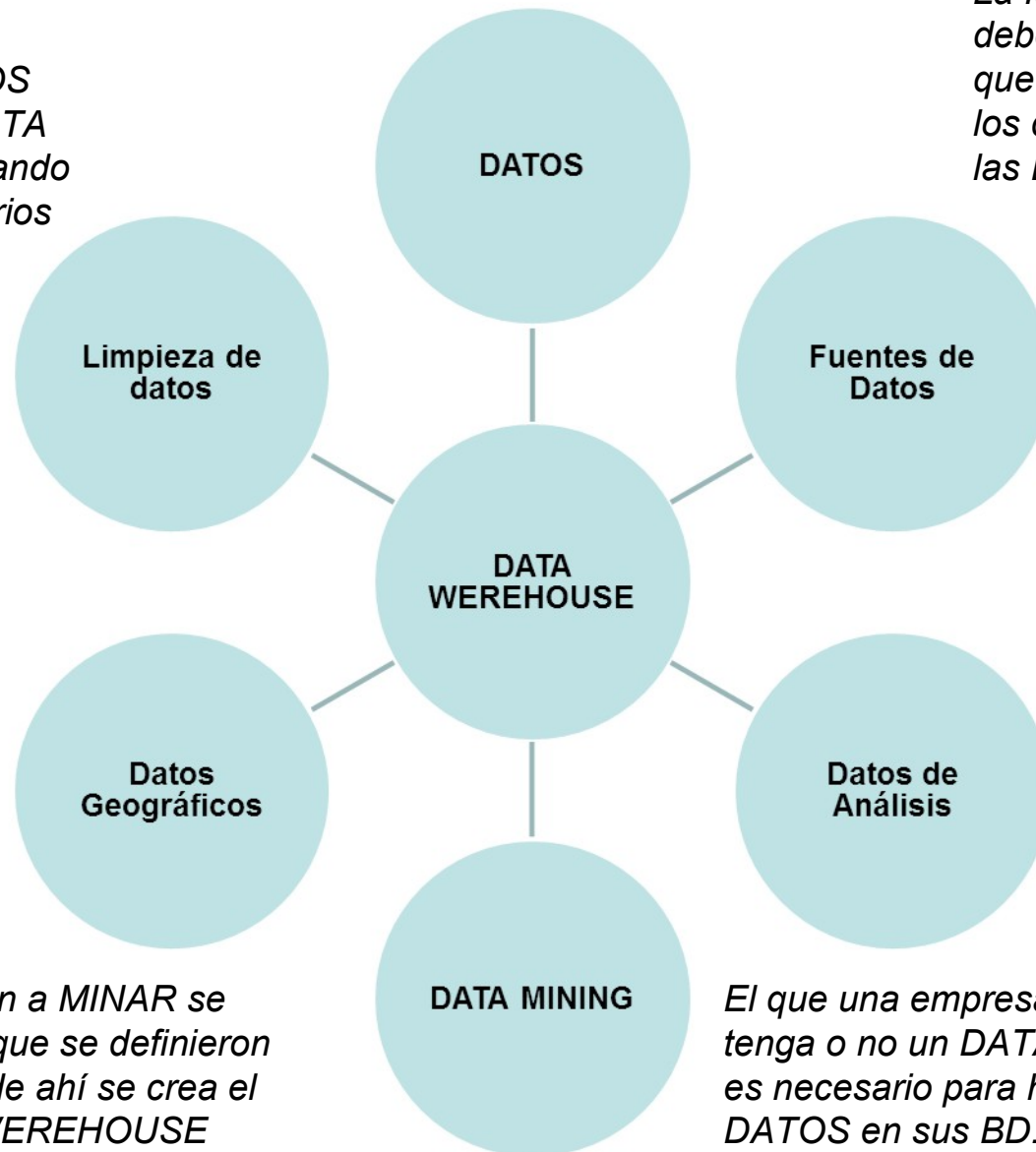




# Introducción a Data Mining

*Se debe de hacer  
LIMPIEZA DE DATOS  
antes de crear el DATA  
WERHOUSE eliminando  
los datos no necesarios*

*La MINERIA DE DATOS  
debe ser mas LOGICA  
que FISICA (VISTAS) de  
los datos contenidos en  
las BD's*



*La BD deben de  
tener soporte  
para las  
demandas de  
minería.*

*La BD deben de  
tener soporte  
para consultas  
OLAP y OLTP*

*Los datos que se van a MINAR se  
obtienen de las BD que se definieron  
como necesarias y de ahí se crea el  
depositorio DATA WERHOUSE*

*El que una empresa u organización  
tenga o no un DATA WERHOUSE no  
es necesario para hacer MINERIA DE  
DATOS en sus BD.*



# Introducción a Data Mining

## DATAWAREHOUSE

Un Datawarehouse es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta.

La creación de un datawarehouse representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implantar una solución completa y fiable de **BUSINESS INTELLIGENCE**.

**BUSINESS INTELLIGENCE** es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios. La inteligencia de negocio actúa como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva, que no es otra que proporcionar **información privilegiada para responder a los problemas de negocio**



# Introducción a Data Mining

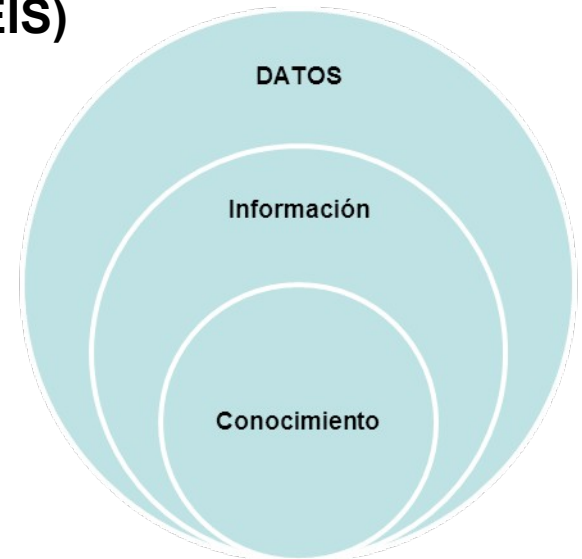
## DATAWAREHOUSE

Los principales productos de **BUSINESS INTELLIGENCE** que existen hoy en día son:

- \* Cuadros de Mando Integrales (CMI)
- \* Sistemas de Soporte a la Decisión (DSS)
- \* Sistemas de Información Ejecutiva (EIS)

Por otro lado, los principales componentes de orígenes de datos en el Business Intelligence que existen en la actualidad son:

- \* DATAMART
- DATAWAREHOUSE



# Introducción a Data Mining

## REFERENCIAS

- IBM, Portal de discusión

[https://www.ibm.com/support/knowledgecenter/es/SSEPGG\\_9.7.0/com.ibm.im.overview.doc/c\\_dm\\_goals.html](https://www.ibm.com/support/knowledgecenter/es/SSEPGG_9.7.0/com.ibm.im.overview.doc/c_dm_goals.html)

- WIKIPEDIA

[https://es.wikipedia.org/wiki/Miner%C3%ADa\\_de\\_datos](https://es.wikipedia.org/wiki/Miner%C3%ADa_de_datos)

- MICROSOFT, portal de documentos y bases de conocimientos

<https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/data-mining-concepts>

- Portal SINEXUS

[http://www.sinnexus.com/business\\_intelligence/](http://www.sinnexus.com/business_intelligence/)

- Material recopilado en las cátedras de Dr. Urube Agundis del 06 al 08/02/2018.