

## PARTE I: INTRODUCCIÓN

**José Hernández Orallo**

jorallo@dsic.upv.es

Departamento de Sistemas Informáticos y Computación  
Universidad Politécnica de Valencia

## Temario

### 1. Introducción

- 1.1. Finalidades y Evolución de los Sistemas de Información.
- 1.2. Herramientas para la Toma de Decisiones: diferencias e interrelación.
- 1.3. Almacenes de Datos, OLAP y Minería de Datos: definición e interrelación.

### 2. Almacenes de Datos

- 2.1. Introducción a los Almacenes de Datos.
- 2.2. Arquitectura de un Sistema de Almacén de Datos.
- 2.3. Explotación de un Almacén de Datos: Herramientas OLAP.
- 2.4. Sistemas ROLAP y MOLAP.
- 2.5. Carga y Mantenimiento de un Almacén de Datos.
- 2.6. Diseño de un Almacén de Datos.
- 2.7. Líneas de Investigación Abiertas.

### 3. Minería de Datos

- 3.1. Introducción a la Minería de Datos (DM)
- 3.2. El proceso de KDD
- 3.3. Técnicas de Minería de Datos
- 3.4. Web Mining
- 3.5. Líneas de Investigación Abiertas

2

## Objetivos Parte I

- Distinguir los sistemas de información para la gestión y los sistemas de información para la toma de decisiones.
- Conocer la evolución de las herramientas para el análisis de sistemas de información para la toma de decisiones.
- Distinguir las propiedades y finalidades de las diferentes herramientas DSS (soporte a la toma de decisiones): EIS, OLAP, consultas & informes, minería de datos.

## Finalidad de los Sistemas de Información

La información *reduce nuestra incertidumbre* (sobre algún aspecto de la realidad) y, por tanto, nos permite tomar mejores decisiones

4

## Finalidad de los Sistemas de Información

- *Inicialmente* la finalidad de los sistemas de información era recopilar información sobre un parcela del mundo para ayudar en la toma de decisiones:
  - recuentos de cereales en Babilonia, de cacao por los pipiles,
  - censos civiles y militares romanos o chinos,
  - libros contables de árabes o sefardies,
  - ...
- *Actualmente*, con la informatización de las organizaciones y la aparición de aplicaciones software operacionales sobre el sistema de información, la finalidad *principal* de los sistemas de información es dar soporte a los procesos básicos de la organización (ventas, producción, personal...).

5

## Interés Renovado por la Finalidad “Prístina”

Una vez satisfecha la necesidad de tener un soporte informático para los procesos básicos de la organización (**sistemas de información para la gestión**).



Las organizaciones exigen nuevas prestaciones de los sistemas de información (**sistemas de información para la toma de decisiones**).

6

## Evolución

- 60's: Informes *batch*:
  - la información es difícil de encontrar y analizar, poco flexible, se necesita reprogramar cada petición.
- 70's: Primeros DSS (*Decision Support Systems*) y EIS (*Executive Information Systems*):
  - basados en terminal, no integrados con el resto de herramientas.
- 80's: Acceso a datos y herramientas de análisis integradas (conocidas como *intelligent business tools*):
  - Herramientas de consultas e informes, hojas de cálculo, interfaces gráficos e integrados, fáciles de usar.
  - Acceden a las bases de datos operacionales ("killer queries").
- 90's: Almacenes de Datos y herramientas OLAP.
- 00's: Herramientas de Minería de Datos y Simulación.

7

## Herramientas para la Toma de Decisiones

Han aparecido diferentes herramientas de negocio o DSS que *coexisten*: EIS, OLAP, consultas & informes, minería de datos, ...

- ¿Cuál es la diferencia entre EIS y OLAP?
- ¿Cuál es la diferencia entre "informes avanzados" y OLAP?
- ¿Cuál es la diferencia entre OLAP y Minería de Datos?
- ¿Qué interrelaciones existen entre todas estas herramientas?

8

## Herramientas para la Toma de Decisiones

¿Cuál es la diferencia entre EIS y OLAP?

- Un EIS (*Executive Information System*) es un sistema de información y un conjunto de herramientas asociadas:
  - Proporciona a los directivos acceso a la información de estado y sus actividades de gestión.
  - Está especializado en analizar el estado diario de la organización (mediante indicadores clave) para informar rápidamente sobre *cambios* a los directivos.
  - La información solicitada suele ser, en gran medida, numérica (*ventas semanales, nivel de stocks, balances parciales, etc.*) y representada de forma gráfica al estilo de las hojas de cálculo.
- Las herramientas OLAP (*On-Line Analytical Processing*) son más genéricas:
  - Funcionan sobre un sistema de información (transaccional o almacén de datos)
  - Permiten realizar agregaciones y combinaciones de los datos de maneras mucho más complejas y ambiciosas, con objetivos de análisis más estratégicos.

9

## Herramientas para la Toma de Decisiones

¿Cuál es la diferencia entre "informes avanzados" y OLAP?

- Los sistemas de informes o consultas avanzadas:
  - están basados, generalmente, en sistemas *relacionales u objeto-relacionales*,
  - utilizan los operadores clásicos: concatenación, proyección, selección, agrupamiento, ... (en SQL y extensiones).
  - el resultado se presenta de una manera tabular.
- Las herramientas OLAP
  - Están basadas, generalmente, en sistemas o *interfaces multidimensionales*,
  - Utilizando operadores específicos (además de los clásicos): *drill, roll, pivot, slice & dice, ...*
  - El resultado se presenta de una manera matricial o híbrida.

10

## Herramientas para la Toma de Decisiones

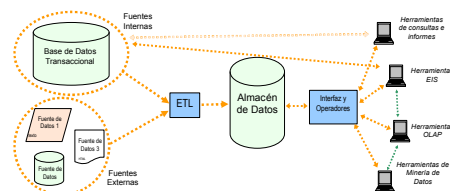
¿Cuál es la diferencia entre OLAP y minería de datos?

- Las herramientas OLAP
  - proporcionan facilidades para "manejar" y "transformar" los datos.
  - **producen otros "datos"** (más agregados, combinados).
  - ayudan a analizar los datos porque producen *diferentes vistas* de los mismos.
- Las herramientas de Minería de Datos:
  - son muy variadas: permiten "extraer" patrones, modelos, descubrir relaciones, regularidades, tendencias, etc.
  - **producen "reglas" o "patrones" ("conocimiento")**.

11

## Herramientas para la Toma de Decisiones

¿Qué interrelaciones existen entre todas estas herramientas?



- La aparición de algunas de ellas han hecho cambiar la manera de trabajar de otras herramientas.

12

## Almacenes de Datos

El almacén de datos es ahora el “**sistema de información central**” en todo este proceso.

Un almacén de datos es una colección de datos:

- orientada a un dominio
- integrada
- no volátil
- variante en el tiempo

para ayudar en la toma de decisiones [Immon 1992, 1996]

13

## Almacenes de Datos

Actualmente,

Los almacenes de datos y las técnicas OLAP son las maneras más efectivas y tecnológicamente más avanzadas para **integrar, transformar y combinar los datos para facilitar** al usuario o a otros sistemas **el análisis de la información**.

- La tecnología OLAP generalmente se asocia a los almacenes de datos, aunque:
  - Podemos tener Almacenes de Datos sin OLAP y viceversa.

14

## Minería de Datos

La Minería de Datos es un conjunto de técnicas de análisis de datos que permiten:

- Extraer patrones, tendencias y regularidades para **describir** y comprender mejor los datos.
- Extraer patrones y tendencias para **predecir** comportamientos futuros.

Debido al gran volumen de datos este análisis ya **no puede ser manual** (ni incluso facilitado por herramientas de almacenes de datos y OLAP) sino que **ha de ser (semi-)automático**.

15

## Minería de Datos

La Minería de Datos se diferencia claramente del resto de herramientas en el sentido de que:

- **no transforma y facilita el acceso a la información para que el usuario la analice más fácilmente.**

la minería de datos “analiza” los datos

16

## Minería de Datos

- La minería de datos es sólo una etapa del **proceso de extracción de conocimiento a partir de datos**.
- Este proceso consta de varias fases:
  - Preparación de Datos (selección, limpieza, y transformación), Minería de Datos, Evaluación, Difusión y Uso de Modelos.
- incorpora muy diferentes técnicas
  - árboles de decisión, regresión lineal, redes neuronales artificiales, técnicas bayesianas, máquinas de soporte vectorial, etc.
- de campos diversos:
  - aprendizaje automático e I.A., estadística, bases de datos, ...
- aborda una tipología variada de problemas:
  - clasificación, categorización, estimación/regresión, agrupamiento, ...

17

## Almacenes de Datos y Minería de Datos

¿Es necesario tener almacenes de datos para realizar minería de datos?

- Los almacenes de datos no son *imprescindibles* para hacer extracción de conocimiento a partir de datos.
  - se puede hacer minería de datos sobre un simple fichero de datos.
- Las *ventajas* de organizar un almacén de datos para realizar minería de datos se amortizan sobradamente a medio y largo plazo cuando:
  - tenemos grandes volúmenes de datos, o
  - éstos aumentan con el tiempo, o
  - provienen de fuentes heterogéneas o
  - se van a combinar de maneras arbitrarias y no predefinidas.

18