

Universidad Autónoma del Estado de Hidalgo

Escuela Superior Huejutla





Área Académica: Sistemas Computacionales

Tema: Introducción a almacén de datos

Profesor: Mtro Felipe de Jesús Núñez Cárdenas

Periodo: Agosto Noviembre 2011

Keywords

Almacén de Datos, Datawarehouse, Arquitectura





Tema: Introducción a almacén de datos

Abstract

Datawarehouses are tools that help us analyze information and whose main features are information-oriented, integrate, time-varying and non-volatile.

Keywords:

Almacen de Datos, Datawarehouse, Arquitectura





Introducción a los Almacenes de Datos

OBJETIVO:

Análisis de Datos para el Soporte en la Toma de Decisiones.

- Generalmente, **la información** que se quiere investigar sobre un cierto dominio de la organización **se encuentra en bases de datos y otras fuentes muy diversas**, tanto internas como externas.
- Muchas de estas fuentes son las que se utilizan para el trabajo diario (**bases de datos operacionales**).





Introducción a los Almacenes de Datos

- Sobre estas mismas bases de datos de trabajo ya se puede extraer conocimiento (visión tradicional).
- Uso de la base de datos transaccional para varios cometidos:
 - Se mantiene el trabajo transaccional diario de los sistemas de información originales (conocido como **OLTP**, ***On-Line Transactional Processing***).
 - Se hace análisis de los datos en tiempo real sobre la misma base de datos (conocido como **OLAP**, ***On-Line Analytical Processing***).





Introducción a los Almacenes de Datos

- Uso de la base de datos transaccional para varios cometidos:
- **PROBLEMAS:**
 - perturba el trabajo transaccional diario de los sistemas de información originales (“*killer queries*”). Se debe hacer por la noche o en fines de semana.
 - la base de datos está diseñada para el trabajo transaccional, no para el análisis de los datos. Generalmente no puede ser en tiempo real (era AP pero no OLAP).





Introducción a los Almacenes de Datos

- Se desea operar eficientemente con esos datos...
 - los costes de almacenamiento masivo y conectividad se han reducido drásticamente en los últimos años,
- parece razonable recoger los datos (información histórica) en **un sistema separado y específico.**

NACE EL DATA-WAREHOUSING

Datawarehouses (Almacenes o Bodegas de Datos)





Introducción a los Almacenes de Datos

Almacenes de Datos (AD)

(data warehouse)



disponer de Sistemas de Información de apoyo a la toma de decisiones*



disponer de *bases de datos* que permitan *extraer conocimiento* de la información histórica almacenada en la organización

objetivos

análisis de la organización

previsiones de evolución

diseño de estrategias

* DSS: Decision Support Systems



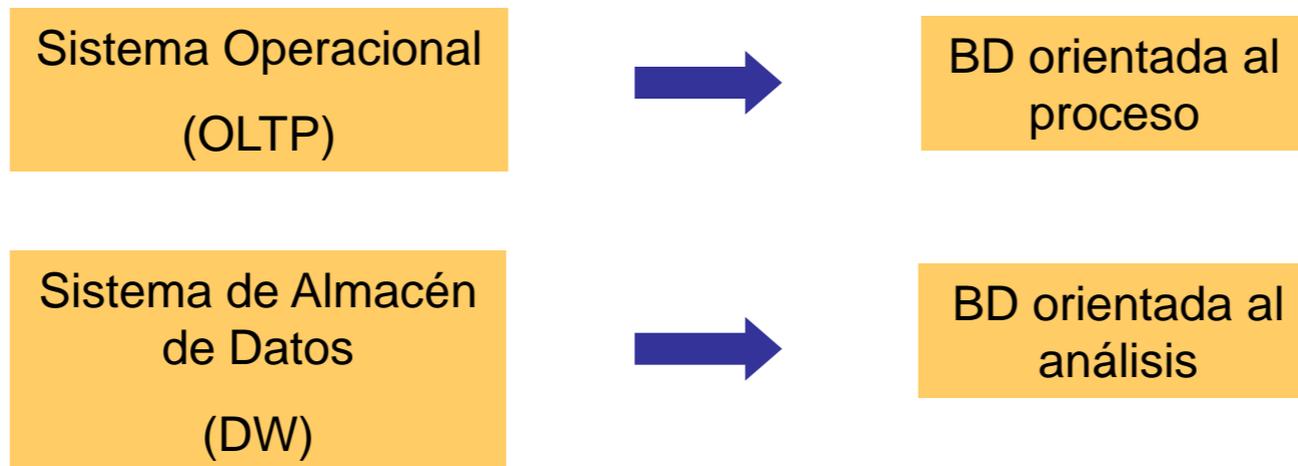


Introducción a los Almacenes de Datos

Almacenes de datos



Base de Datos diseñada con un objetivo de explotación distinto que el de las bases de datos de los sistemas operacionales.





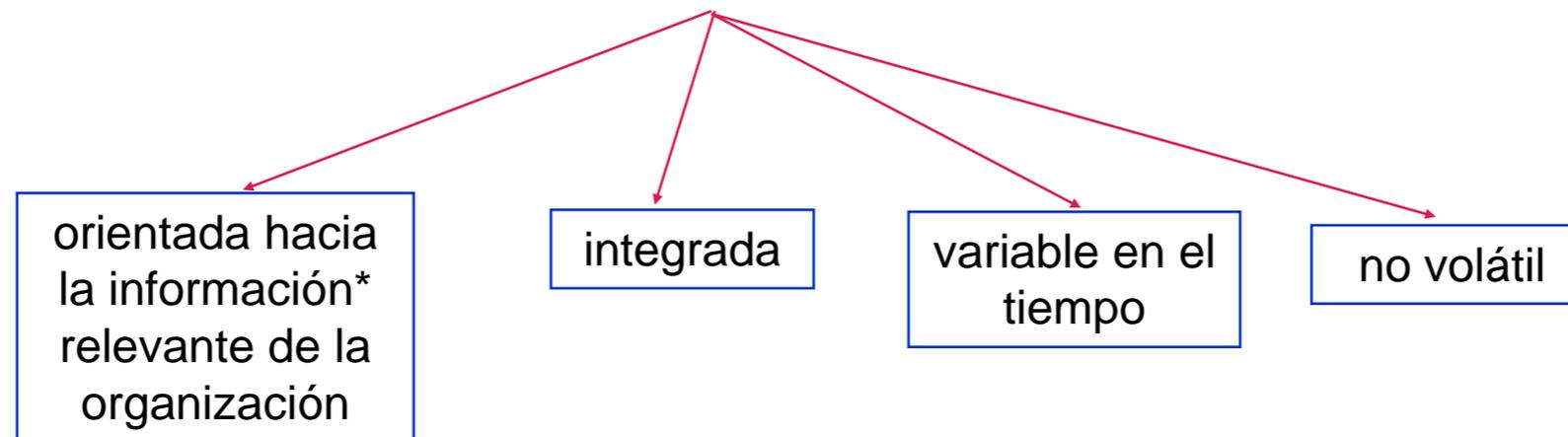
Introducción a los Almacenes de Datos

Almacenes de Datos



colección de datos diseñada para dar apoyo a los procesos de toma de decisiones

características



** subject oriented, not process oriented*



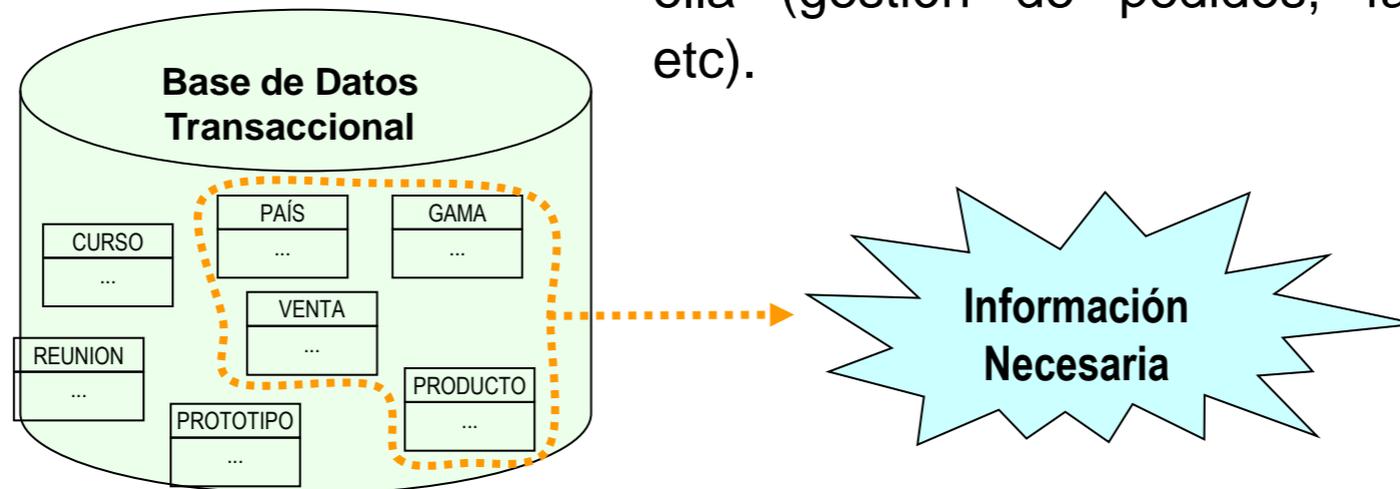


Introducción a los Almacenes de Datos

AD: Orientado hacia la información relevante de la organización



se diseña para consultar eficientemente información relativa a las actividades (ventas, compras, producción, ...) básicas de la organización, no para soportar los procesos que se realizan en ella (gestión de pedidos, facturación, etc).



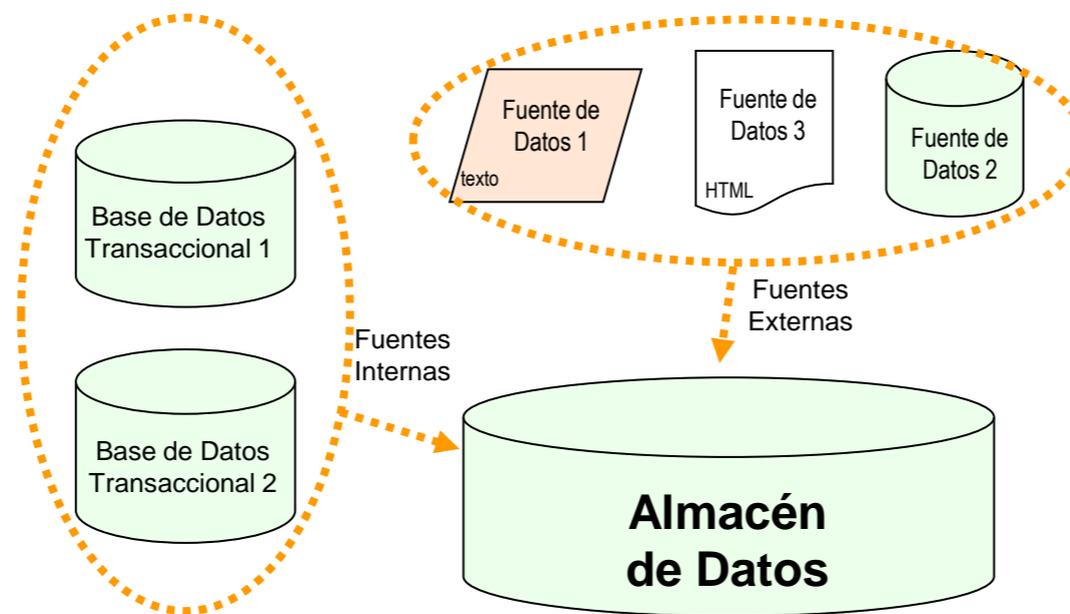


Introducción a los Almacenes de Datos

**AD:
Integrado**



integra datos recogidos de diferentes sistemas operacionales de la organización (y/o fuentes externas).





Introducción a los Almacenes de Datos

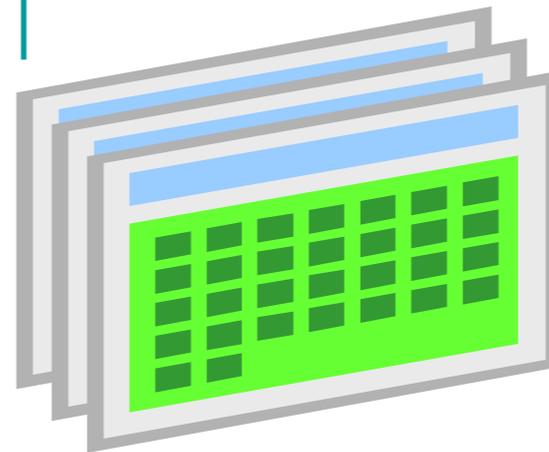
AD: Variable
en el tiempo



los datos son relativos a un periodo de tiempo y deben ser incrementados periódicamente.

Los datos son almacenados como fotos (snapshots) correspondientes a periodos de tiempo.

Tiempo	Datos
01/2003	Datos de Enero
02/2003	Datos de Febrero
03/2003	Datos de Marzo



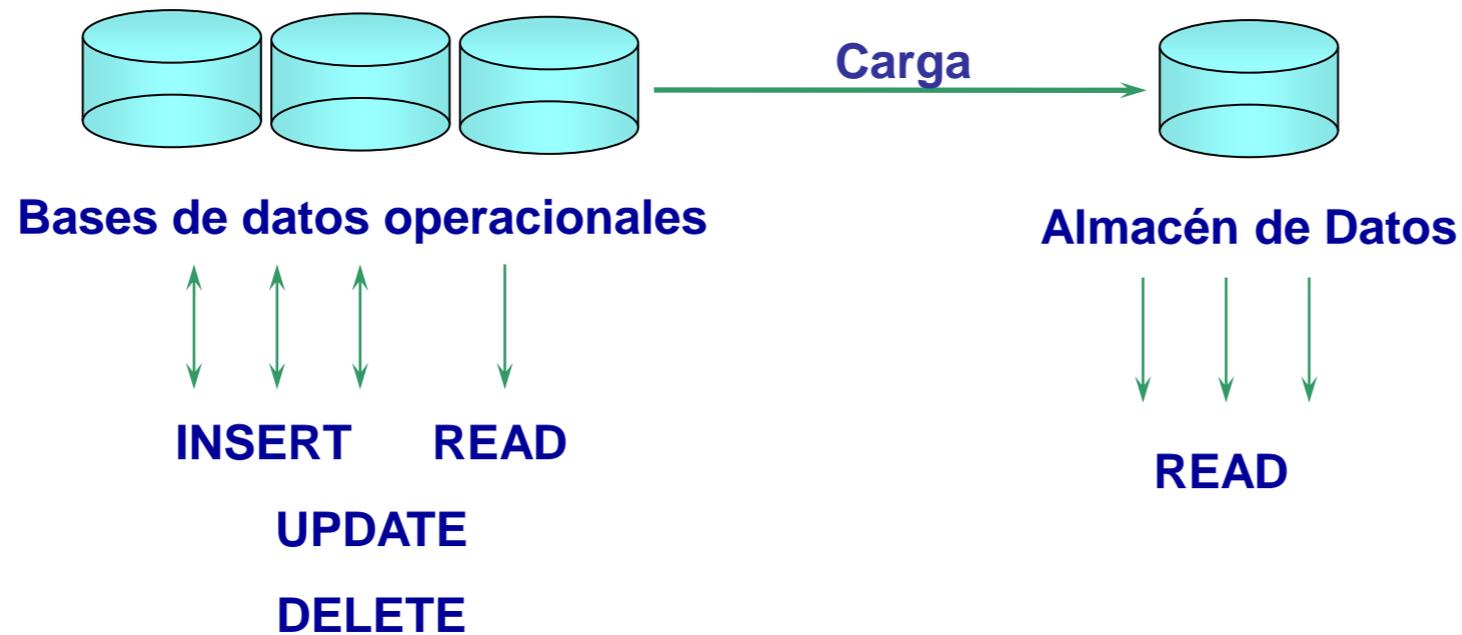


Introducción a los Almacenes de Datos

AD: No volátil



los datos almacenados no son actualizados, sólo son incrementados.

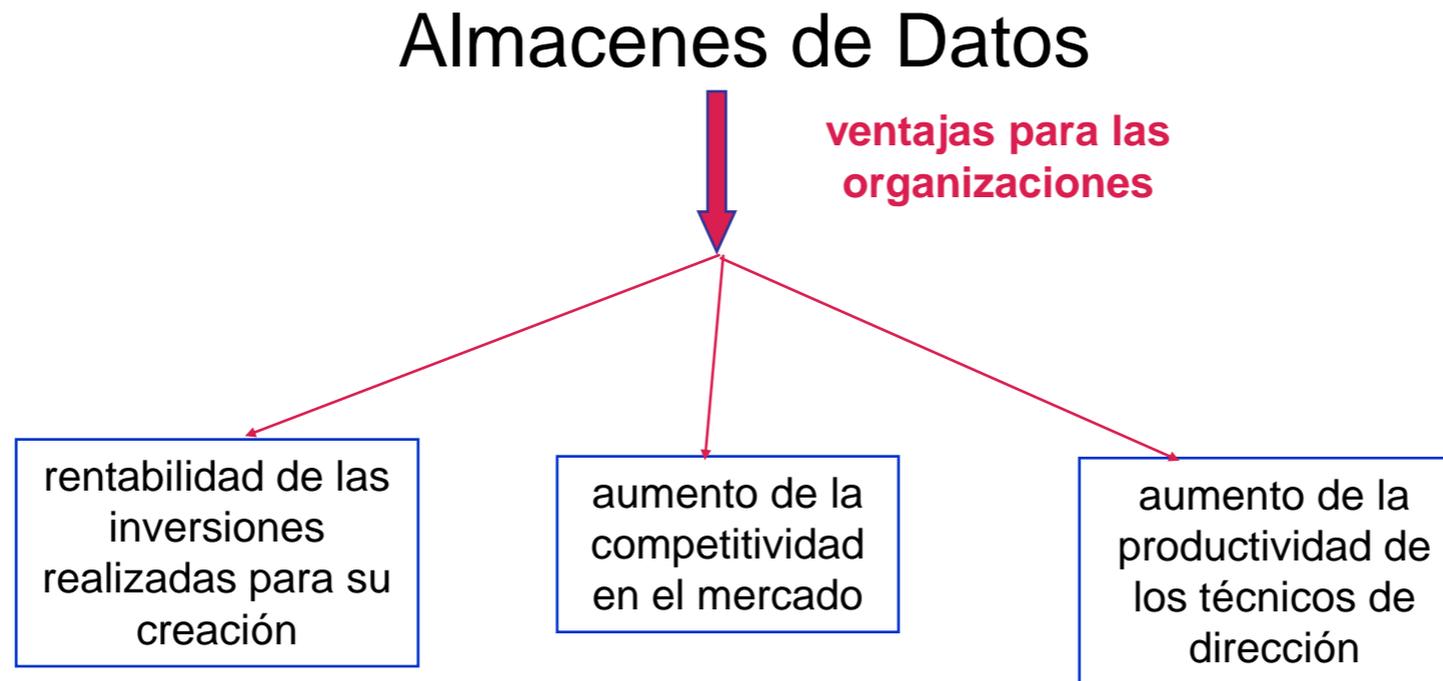


El periodo de tiempo cubierto por un AD varía entre 2 y 10 años.



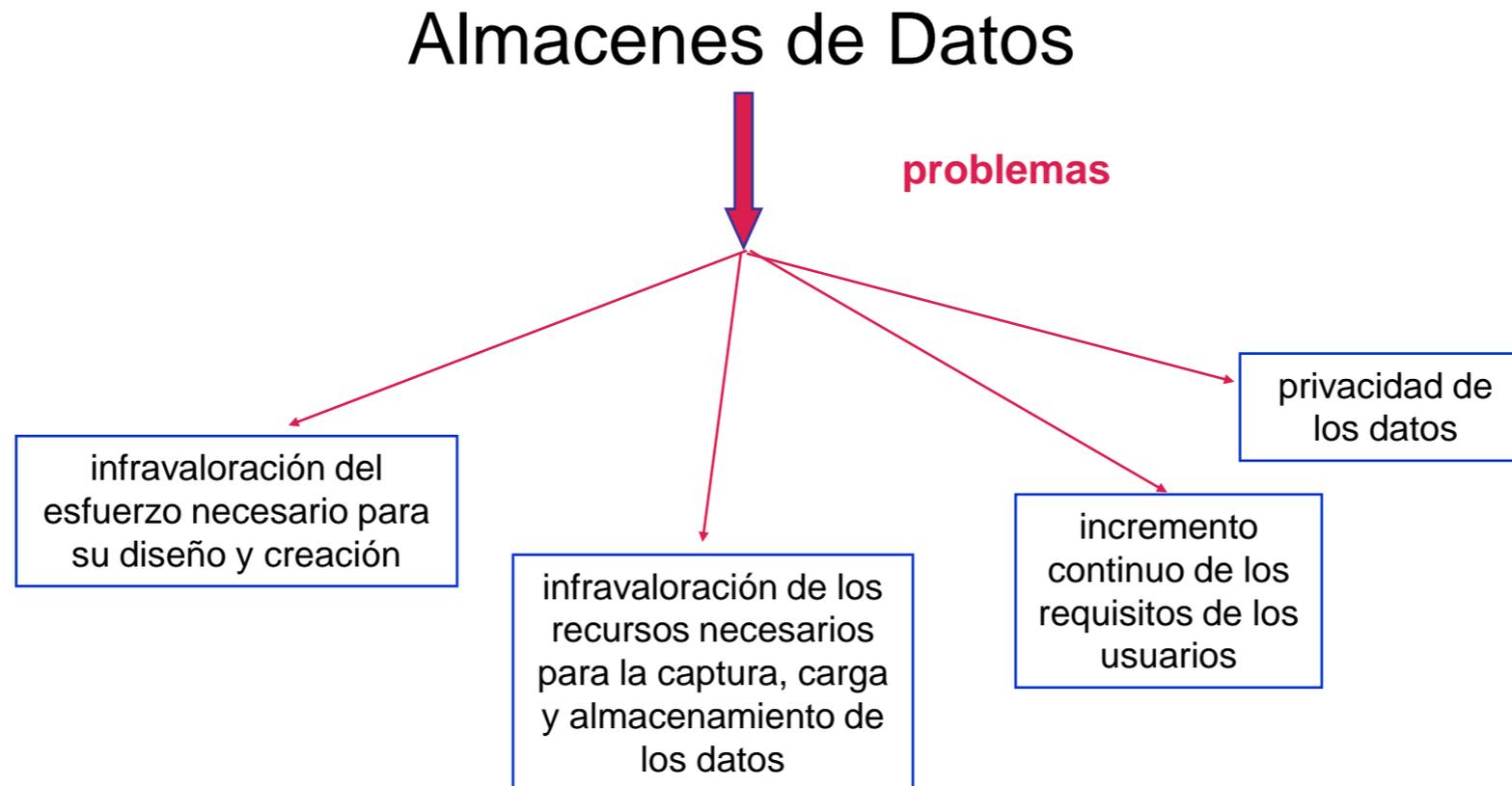


Introducción a los Almacenes de Datos





Introducción a los Almacenes de Datos





Introducción a los Almacenes de Datos

Sistema Operacional (OLTP)

- almacena datos actuales
- almacena datos de detalle
- bases de datos medianas (100Mb-1Gb)
- los datos son dinámicos (actualizables)
- los procesos (transacciones) son repetitivos
- el número de transacciones es elevado
- tiempo de respuesta pequeño (segundos)
- dedicado al procesamiento de transacciones
- orientado a los procesos de la organización
 - soporta decisiones diarias
 - sirve a muchos usuarios (administrativos)

Almacén de datos (DW)

- almacena datos históricos
- almacena datos de detalle y datos agregados a distintos niveles
- bases de datos grandes (100Gb-1Tb)
- los datos son estáticos
- los procesos no son previsibles
- el número de transacciones es bajo o medio
- tiempo de respuesta variable (segundos-horas)
- dedicado al análisis de datos
- orientado a la información relevante
- soporta decisiones estratégicas
 - sirve a técnicos de dirección





Bibliografía

- Hand, D.J.; Mannila, H. and Smyth, P. “Principles of Data Mining”, The MIT Press, 2000.
- Hernández, J.; Ramírez, M.J.; Ferri, C. “Introducción a la Minería de Datos” Pearson Prentice Hall, 2004.
- Kosala, R.; Blockeel, H. “Web Mining Research: A Survey” ACM SIGKDD Explorations, Newsletter of the ACM SIG on Knowledge Discovery and Data Mining, June 2000, Vol. 2, nº1, pp. 1-15.
- Mena, Jesus “Data Mining Your Website”, Digital Press, July 1999.
- Mitchell, T.M. “Machine Learning” McGraw-Hill 1997.
- Pyle, D. “Data Preparation for Data Mining” Morgan Kaufmann, Harcourt Intl., 1999.
- Thuraisingham, B. “Data Mining. Technologies, Techniques, Tools, and Trends”, CRC Press, 1999.
- Witten, I.H.; Frank, E. “Tools for Data Mining”, Morgan Kaufmann, 1999.
- Wong, P. C. “Visual Data Mining”, Special Issue of *IEEE Computer Graphics and Applications*, Sep/ Oct 1999, pp. 20- 46.
- Material extraído del Análisis y Extracción de Conocimiento en Sistemas de Información: Datawarehouse y Datamining de **José Hernández Orallo**, **Universidad Politécnica de Valencia**

