

## Base de Datos II

Unidad 6 – Apoyo a la toma de Decisiones  
Recuperación de Información valiosa

### Objetivos

- ▶ Necesidades especiales para la toma de decisiones
- ▶ Data-Warehousing.
- ▶ Minería de datos: Data-Mining.



## **“Nuevos Tiempos” en las organizaciones**

---

- Tres fuerzas impulsan a las compañías a producir un cambio.
  - Las tres Ces!!:
    - Clientes
    - Competencia
    - Cambio
- 



## **Los Clientes asumen el mando**

---

- Cambia la fuerza dominante en la relación Vendedor-Cliente.
  - Antes: Los clientes se adaptaban. La oferta era poca.
  - Hoy: La empresa se adapta al cliente. La oferta es mucha.
  - Los Clientes exigen PERSONALIZACION.
  - Mayor diversificación.
  - La TECNOLOGÍA es la gran impulsora del cambio.
- 



## **La Competencia se intensifica**

---

- Se diversifica sobre bases competitivas totalmente distintas.
  - No existen territorios protegidos de competencia extranjera.
  - Un competidor eficiente sube el umbral competitivo para los demás.
  - Nuevos competidores.
  - La TECNOLOGÍA cambia la naturaleza de la competencia.
- 

## **El Cambio se vuelve constante**

---

- General y permanente.
  - La rapidez del cambio TECNOLÓGICO promueve la innovación.
  - Cortos ciclos de vida de los productos.
  - Disminuye el tiempo disponible para innovar.
-

## ¿Qué sucede?

---

- La división del trabajo y la especialización de las tareas estratificaron los procesos separando aquellas actividades relacionadas a la producción de aquellas asociadas a la gestión.
  - Podemos categorizar entonces a los procesos en algún punto en un continuo entre dos extremos:
    - Los completamente estructurados.
    - Los completamente no estructurados.
- 



## ¿Qué sucede?

---

- Paradójicamente, mientras la evolución de las organizaciones se mueve hacia ambientes menos estructurados, los sistemas han evolucionado en general dentro de los ambientes estructurados.
  - El volumen de información empresarial se encuentra en un crecimiento constante. Muchos datos están dejando de ser estudiados debido a la incapacidad de los sistemas actuales para procesarlos.
- 



## ¿Qué sucede?

- Se han generado y se generan muchos datos. El desafío es encontrar la información que hay oculta y darle un significado.
- La tecnología hoy dejó de ser una facilitadora para ser una posibilitadora de la administración de la información de los procesos.
- La tecnología de la información está transformando la forma de hacer negocios y los negocios mismos.



## ¿Qué sucede?

- Las grandes innovaciones de la humanidad fueron tecnologías genéricas. Han producido éxito en la medida que fueron aplicadas exitosamente. Requirieron de siglos para ofrecer su máxima utilidad.
- Hoy indudablemente LA innovación pasa por la información en todas sus formas.
- Juegan entonces un papel preponderante los desarrollos que se hagan para aplicar esta innovación.



## ¿Qué sucede?

- Seguramente no va a tomar siglos obtener los máximos beneficios de estas tecnologías, pero todavía no sabemos el impacto final que va a tener.
- Se trata de estar atento y al día.



## "Data Warehouses" - Motivación

- En la mayor parte de las organizaciones productivas y de servicios **las bases de datos operacionales han sido diseñadas para cubrir las necesidades de procesos eminentemente transaccionales.**
- Los sistemas transaccionales que operan sobre las bases de datos realizan consultas o "queries" sencillas, que pueden ser eficientemente respondidas por los administradores de bases de datos.
- Desafortunadamente, no importa cuán eficientemente se hayan diseñado las **bases de datos de una organización, generalmente no satisfacen las necesidades de información de los sistemas de soporte de decisión (SSD) y los sistemas de soporte para ejecutivos (ESS)**

## "Data Warehouses" - Motivación

---

- Una de las razones fundamentales de esta aseveración es que las bases de datos tradicionales **no poseen información histórica y la estabilidad de datos** que es necesaria para llevar a cabo el tipo de análisis que realizan quienes toman decisiones.
  - **¿QUÉ TIPOS DE ANÁLISIS REALIZAN QUIENES TOMAN DECISIONES?**
- 

## Ejemplos de Consultas de Nivel Estratégico y Táctico

---

- ¿Cómo afectó las ventas del producto X la promoción navideña, en comparación con la misma promoción del año anterior? ¿Fue la performance distinta en diferentes regiones?
  - ¿Cómo influyó en la tasa de accidentes industriales el nuevo plan de capacitación de operarios? ¿El impacto fue diferente en las distintas plantas de la empresa?
  - ¿Cuál es la relación entre la cantidad de material "scrap" generado y el número de diferentes productos manufacturados en un mes de operación? ¿Existe relación entre la cantidad de "scrap" generado y el rendimiento económico de un determinado proceso productivo?
-

### **Por qué las BD tradicionales son inapropiadas para soportar este tipo de consultas?**

- Es necesario disponer de información histórica.
- Se requiere combinar datos extraídos de bases de datos diferentes, a menudo implementadas sobre distintos administradores tales como Oracle, Access, Informix, SQL Server, etc.
- Las bases de datos operacionales fueron concebidas para satisfacer requerimientos de seguridad y eficiencia de sistemas transaccionales y si se las emplea para soportar las actividades de sistemas de soporte de decisión, no sólo no proveerán respuestas satisfactorias sino que además degradarán la performance de los sistemas transaccionales a los que naturalmente deben soportar.



### **¿Cómo apoyar la actividad de los Sistemas de Soporte de Decisión?**

- Históricamente, los sistemas de soporte de decisión se emplearon en forma muy limitada, ya que no se disponía de la información necesaria en un tiempo razonable.
- Para salvar este inconveniente se propuso "congelar" determinados extractos para tareas de análisis. Estos extractos contenían información acerca de un conjunto de entidades seleccionadas en un determinado punto en el tiempo.
- Mientras el empleo de una serie de extractos (fotos de la base de datos en el tiempo) resultaba más eficiente que utilizar la base de datos transaccional directamente, no proveía una solución integral y carecía de la flexibilidad necesaria para responder a preguntas "ad-hoc".





## Data Warehouses (DW)

- Se concibieron para soportar las necesidades de información de los SSD y ESS, es decir dar apoyo a los niveles gerenciales medios y altos de una empresa.
- Una DW es un sistema administrador de bases de datos que existe separadamente de los DBMS que soportan las actividades operacionales, pero que se integra con los mismos.
- Los datos almacenados en una DW **son no volátiles, o son mucho menos volátiles** que en una base de datos tradicional, y en consecuencia permiten la realización de una variedad de análisis. Por ejemplo, mientras en una base de datos operacional se almacena el precio actual de una materia prima, en una DW se mantendrá información histórica del valor de este atributo de las materias primas.



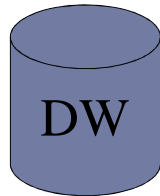
## Data Warehouses (DW)

- Las DW son bases de datos relacionales que soportan una amplia variedad de consultas, que pueden ser formuladas en diferentes formatos.
- Contienen cientos de tablas cuyo diseño se ha optimizado para las consultas típicas.
- Para su creación los datos son tomados de una variedad de fuentes de datos operacionales. Una vez transferidos, estos datos son filtrados para asegurar que los mismos tienen significado, que son consistentes y que poseen el nivel de exactitud adecuado. Luego son cargados en tablas relacionales que soportan una variedad de análisis y consultas y cuya estructura se ha optimizado para satisfacer aquellas que se espera ocurran con mayor frecuencia.



## Data Warehouses (DW)

- Base de Datos para dar soporte a todos los procesos de decisión de la organización.



### Características

- orientación a temas
- integración
- relación con el tiempo
- no volátil



## DW – Orientado a temas

- Es organizado en torno a TEMAS principales
- Excluye datos no usados para tomar decisiones
- Los PROCESOS DE NEGOCIO no son considerados en el diseño del DW.



## DW – Integración

---

### Fuentes de Datos del DW

- Datos Internos
  - Entorno operacional de la organización
  - Legacy Systems, DB operacionales, spreadsheet, etc.
- Datos Externos



## DW – Relación con el tiempo

---

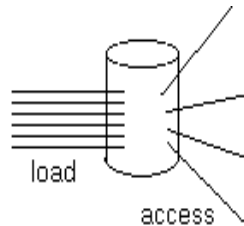
### Fotografía del Dato:

- Horizonte tiempo - 5-10 años
- Clave con atributos de tiempo
- Una vez tomada la fotografía, el registro no se cambia.



## DW – No volátil

- El dato es cargado en el DW y accedido, pero una vez realizada la fotografía del dato, el dato en el DW no se modifica.



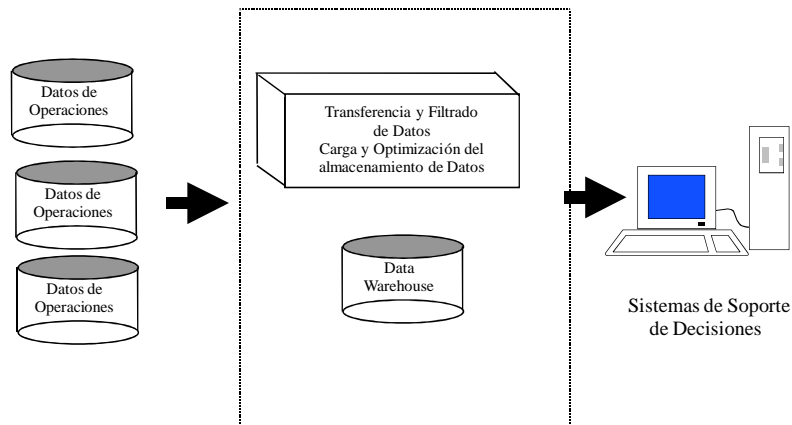
## DW – Consecuencias de la No volátilidad

### Manejo de datos mucho más simple

- Elimina todo el proceso de actualización de datos en-línea:
  - backup y recuperación,
  - transacción e integridad,
  - detección y solución de deadlock.
- Es posible optimizar el acceso a datos



## Representación Esquemática de la Creación y Operación de una Data Warehouse



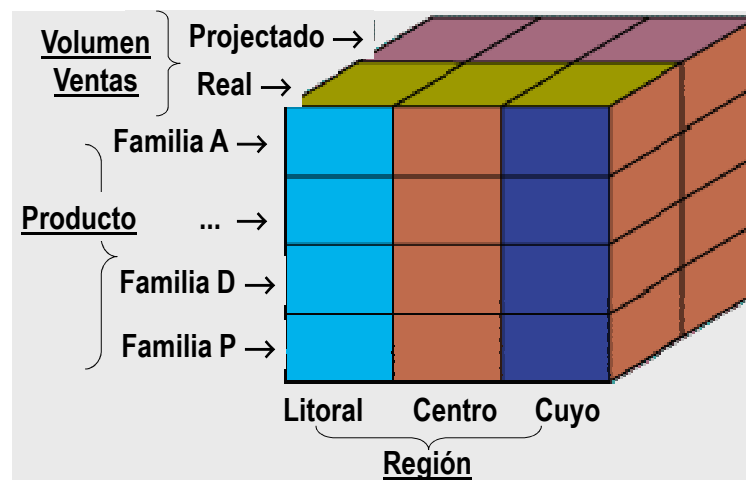
## Creación de una Data Warehouse

- La creación de una DW es un proceso difícil y altamente consumidor de tiempo, por lo que sus costos son generalmente elevados.
- Mientras los procesos de transferencia de datos son sencillos, los procesos de identificación de datos relevantes, su "mezclado" y correcto filtrado son sumamente difíciles.
- Con el propósito de hacer la base de datos más eficiente, los diseñadores enfrentan decisiones tales como sobre qué campos indexar los datos y cuál es el formato de indexado que es más conveniente desde el punto de vista del uso que se hará de los mismos.
- Para la construcción de DWs se realizan numerosas entrevistas que tienen el propósito de definir las necesidades de todos los usuarios y de contemplar todos los posibles usos que podrían darse a los datos.
- Una DW se construye en forma incremental.

## Data Warehouses (DW)

- El diseño de una DW no debe ser estático.
- A medida que la cantidad de información se multiplica en el tiempo, los diseñadores de la DW pueden que necesiten definir nuevas sintaxis, así como otros formatos de consulta que sean más rápidos y sencillos de utilizar.
- De manera similar, a medida que aumenta la carga de información es menester implementar nuevos mecanismos de unión de tablas relacionales y de **búsqueda de la información relevante**.
- Esta última actividad es también conocida como "**data mining**" y se lleva a cabo con la ayuda de una serie de tecnologías, entre las que merecen citarse: **Técnicas de visualización, OLAP (On Line Analytical Processing), Agentes Inteligentes**, entre otras.
- Las técnicas OLAP se emplean para realizar un modelado multidimensional de los datos, soportando la modelización y análisis de grandes volúmenes de datos desde diferentes perspectivas.

## Modelo de Análisis Multidimensional



## Data Mining y Agentes Inteligentes

- A pesar que las DW proveen acceso a información que va a ayudar a los tomadores de decisiones a comprender su ambiente de trabajo, la gran variedad de oportunidades que brindan puede ser contraproducente.
- Puede que los usuarios no experimentados sean sobrepasados por las capacidades de la DW y requieran de herramientas especializadas que los ayude a encontrar la información que ciertamente es de valor.
- Los **agentes inteligentes** pueden ayudar a hallar los datos que mejor se ajustan a un determinado patrón, a encontrar la mejor alternativa entre un conjunto de posibilidades, a identificar los casos que se ajustan a un determinado criterio, etc. También pueden analizar grandes volúmenes de información para hallar patrones y descubrir tendencias.

## Data Mining y Agentes Inteligentes

- Los **agentes inteligentes** son "software" que llevan a cabo tareas específicas, en lugar de los usuarios, actuando a menudo en forma autónoma, es decir realizando la mayoría de sus actividades sin intervención humana.
- No son nuevos, ya que desde hace mucho tiempo se los ha empleado para el monitoreo de las capacidades y del uso que se hace de la CPU, así como para clasificar y priorizar la correspondencia electrónica recibida. Actualmente, los más populares son los agentes vinculados a Internet
- Sin embargo, los agentes inteligentes modernos tienen nuevas y mejores capacidades para **bucear en los datos** ya que en su definición se han incorporado una variedad de herramientas de Inteligencia Artificial (Redes Neuronales, Herramientas para Reconocimiento de Patrones, Razonamiento basado en casos).

## Agentes Inteligentes

- Los agentes inteligentes trabajan en diferentes dominios del conocimiento y con diferentes aplicaciones informáticas (Ayudantes de correo electrónico, personalización de páginas Web, etc.). **Existen agentes inteligentes que operan en conjunción con una DW** y su empleo debe ser programado.
- Pueden ser invocados de diferentes formas:
  - de manera eventual, cuando el usuario lo considera apropiado,
  - periódicamente, de acuerdo a una determinada agenda o a un "schedule" de trabajo,
  - en base a la ocurrencia de eventos predeterminados, tales como el alcance un determinado valor por parte de un indicador, la aplicación de una determinación promoción, etc.



## Agentes Inteligentes

- Muchos agentes inteligentes, basados en diferentes metodologías (estadísticas, redes neuronales, técnicas de componentes principales, etc.), proveen una variedad de opciones para "scanear" la DW y buscar información. Por ejemplo, los usuarios pueden definir filtros basándose en diferentes aspectos, tales como:
  - criterios de calificación,
  - percentiles,
  - información ordinal o basada en "rankings" o jerarquías.





## Posibles Resultados de Actividades de “Data-Mining”

---

- Asociaciones.
  - Secuencias
  - Clasificaciones
  - Agrupaciones o “Clusters”
  - Pronósticos
- 

## Resultados de la Actividad de “Data-Mining”

---

- **Asociaciones:** Se corresponden con eventos vinculados a una situación o a un criterio. Por ejemplo, artículos que la gente suele comprar simultáneamente, o cursos optativos que los alumnos suelen tomar simultáneamente.
  - **Secuencias:** Se corresponden con una serie de eventos que tienen lugar secuencialmente en el tiempo. Por ejemplo, la secuencia de cursos optativos que los buenos alumnos de una carrera toman a lo largo de sus estudios, en un esquema de cursado con créditos; secuencias de eventos que ocurren en un ambiente industrial
-

## Resultados de la Actividad de “Data-Mining”

---

- **Clasificaciones:** Se trabaja sobre un grupo de "individuos" muestrales que ya han sido clasificados (por ejemplo, personas con un ingreso per cápita superior a un determinado monto, o inferior a un valor, alumnos con un promedio de notas superior a un valor umbral, etc.) y se trata de identificar patrones de comportamiento en ellos, es decir valores comunes en determinados atributos.
- 

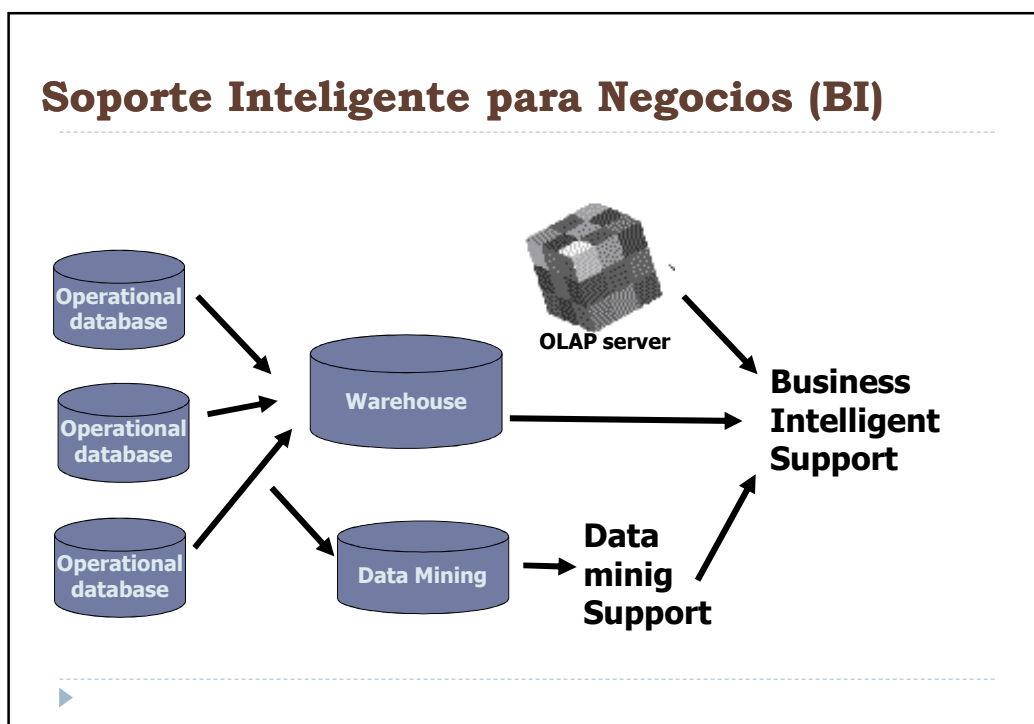
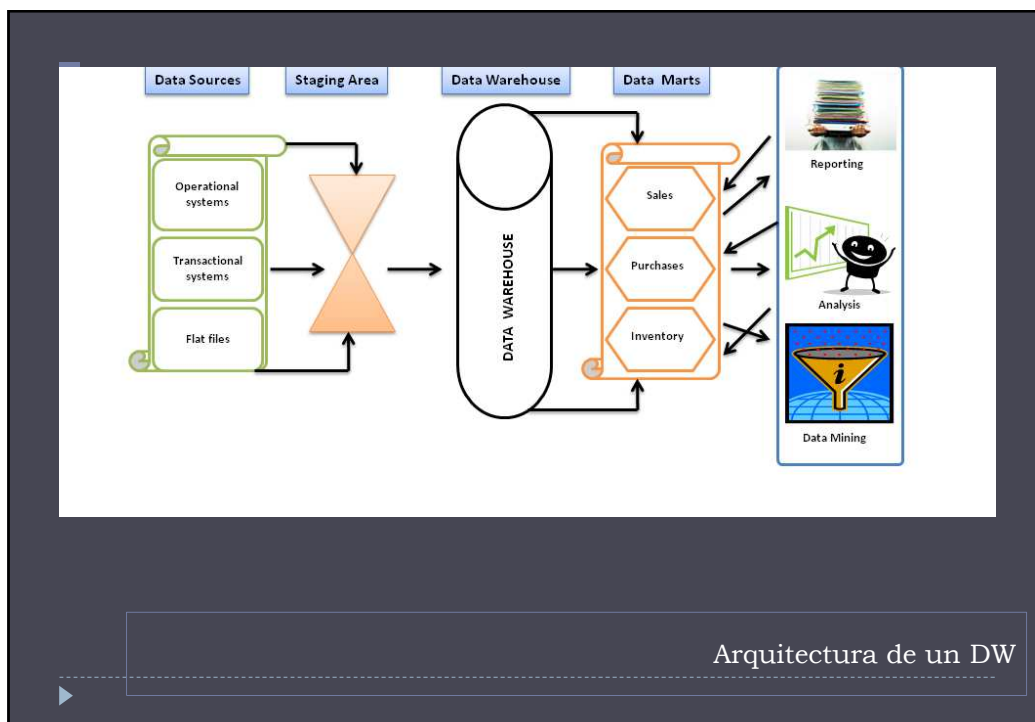


## Resultados de la Actividad de “Data-Mining”

---

- **Agrupaciones o "Clusters":** Un proceso similar al anterior es el de "clustering", en el que se trata de inferir reglas sobre grupos, que se distinguen de otros grupos. La diferencia entre clasificación y "clustering" se basa en el hecho que en el primer caso, las categorías se conocen de antemano (Por ejemplo, las personas de ingresos altos, medios y bajos), mientras que en el caso de "clustering" los agentes inteligentes también identifican los diferentes grupos, es decir deben definir un agrupamiento y encontrar reglas de comportamiento para cada uno de los grupos identificados.
  - **Pronósticos:** Es la más conocida y la que se ha realizado por más tiempo con diferentes tipos de herramientas.
- 





### **Criterios de Evaluación de Productos para “Data-Mining”**

- Qué enfoque emplea la herramienta para modelar los datos?
- Qué tipo de problemas permite abordar ?
- Qué “hardware” y sistema operativo demanda ?
- Requiere un extracto de la DW o puede leer directamente de ella ? En caso afirmativo, sobre qué DBMS debe estar soportada para que pueda hacer la lectura de información ?
- Qué tipo de interfaz de usuario posee para realizar el ingreso y la interpretación de datos ? Es fácil de utilizar y comprender?
- Cuál es el máximo número de variables y filas que soporta ?

### **Criterios de Evaluación de Productos para “Data-Mining”**

- Cuán sensible a “ruidos” es la herramienta ? A qué tipo de “información corrupta” es particularmente sensible ? En qué extensión ?
- Cuál es el tiempo de respuesta promedio requerido para obtener resultados útiles ?
- Qué tipo de resultados puede entregar? (Asociaciones, Secuencias, Clasificaciones, pronósticos, etc.)
- Cuán claros son los resultados?
- Qué tipo de documentación está disponible ?
- Posee ayuda “on-line” ?
- Qué soporte a usuarios provee ?

Es posible apreciar que muchos de los criterios de evaluación son los que se emplean cuando se analizan diferentes herramientas informáticas provistas por empresas de software



## Objetivos

---

- ✓ Necesidades especiales para la toma de decisiones
- ✓ Data-Warehousing.
- ✓ Minería de datos: Data-Mining.

