

# Almacenes de datos: motivación y características principales



---

Tema 1

**Profesor:**

Francisco Rodríguez Gómez



# Indice

---

- Repaso de la tecnología de bases de datos
- Nuevas necesidades de gestión de datos
- El Almacén de datos: primera aproximación
- Diferencias entre sistemas operacionales y sistemas de almacenes de datos



# Indice

---

- Repaso de la tecnología de bases de datos
- Nuevas necesidades de gestión de datos
- El Almacén de datos: primera aproximación
- Diferencias entre sistemas operacionales y sistemas de almacenes de datos

## Repaso de la tecnología de bases de datos

- Evolución de los sistemas de almacenamiento de datos
  - Ficheros
    - Inconsistencia
    - Redundancia
    - Responsabilidad de control en las aplicaciones
  - SGBD (CRUD)
    - Definición, manipulación, y consulta
    - Interfaz usuario-sistema
    - Usuario<->APP Cliente<->APP Servidor<->SGBD
    - Independencia de los datos (separados de los programas)





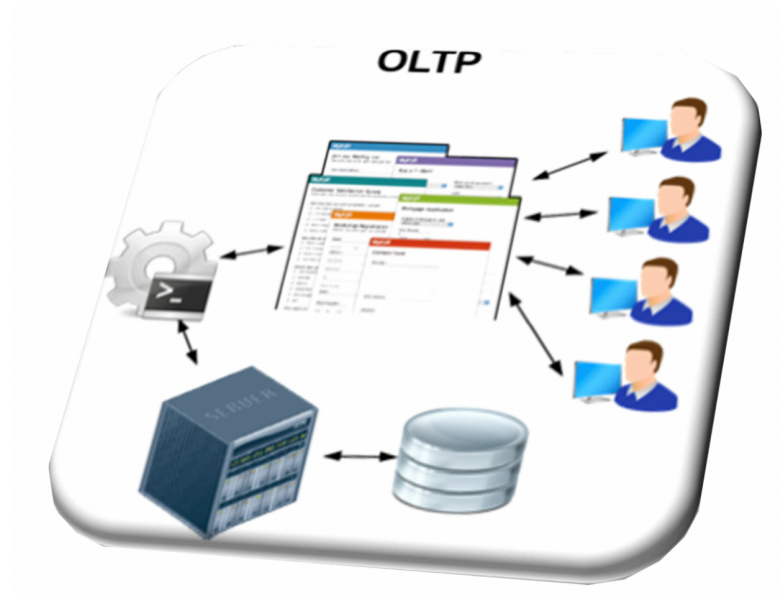
## Repaso de la tecnología de bases de datos

---

- Actualidad
  - SGBD Relacionales (SQL)
    - La mayoría de las de las empresas
    - Eliminación de inconsistencia y redundancia
    - Oracle, SQL Server, MySql, ...
  - SGBD Relacionales (NO SQL)
    - MongoDG, ...

## Repaso de la tecnología de bases de datos

- Sistemas OLTP
  - On-Line Transactional Processing (OLTP)
    - Orientados a las transacciones
      - Compras, ventas, pedidos, clientes, ...





# Indice

---

- Repaso de la tecnología de bases de datos
- Nuevas necesidades de gestión de datos
- El Almacén de datos: primera aproximación
- Diferencias entre sistemas operacionales y sistemas de almacenes de datos



# Indice

---

- Repaso de la tecnología de bases de datos
- Nuevas necesidades de gestión de datos
- El Almacén de datos: primera aproximación
- Diferencias entre sistemas operacionales y sistemas de almacenes de datos





## Nuevas necesidades de gestión de datos

---

- Alta competitividad económica
  - Tecnología y herramientas para la toma de decisiones estratégicas
    - Necesidad de descubrir información oculta en los datos
    - Presentación de información relevante de forma comprensible
    - Respuestas a preguntas de negocio



## Nuevas necesidades de gestión de datos

---

- ¿Ha variado un cliente sus gustos de compra de vehículos? ¿Ha estado comprando el mismo vehículo de soltero que de casado?
- ¿Qué descuento deberíamos ofrecer para incrementar significativamente las ventas?
- El producto de tipo A se ha vendido mucho más que el producto de tipo B. Encontrar patrones de correlación entre estos tipos de productos y extraer similitudes/diferencias
- ¿Cuáles son las tendencias en mis ventas ?
- ¿Qué tipo de cliente me ha estado comprando el BMW 320i durante los últimos 10 años ?
  - ¿Directivo, profesor, trabajador escala básica , etc.?

## Nuevas necesidades de gestión de datos

- Req. 1. Velocidad en las respuestas
  - Los sistemas OLTP no guardan históricos
  - Acceso a distintas fuentes



¿Ha variado un cliente sus gustos de compra de vehículos?

- Req. 2. Respuestas presentadas en un entorno amigable y fácil de usar
  - Entender el tipo de preguntas “estratégicas”



## Nuevas necesidades de gestión de datos

---

- ¿ Son válidos los sistemas OLTP para tales decisiones?
  - Problemas
    - Tiempo de respuesta
    - No entienden preguntas de negocio
    - Falta de herramientas para la extracción y presentación de información relevante para dar respuesta a las preguntas de negocio
    - Los directivos no obtienen resultados en formato comprensible y en tiempos aceptables



# Indice

---

- Repaso de la tecnología de bases de datos
- Nuevas necesidades de gestión de datos
- El Almacén de datos: primera aproximación
- Diferencias entre sistemas operacionales y sistemas de almacenes de datos



# Indice

---

- Repaso de la tecnología de bases de datos
- Nuevas necesidades de gestión de datos
- El Almacén de datos: primera aproximación
- Diferencias entre sistemas operacionales y sistemas de almacenes de datos



## El almacén de datos: primera aproximación

---

- El Almacén de datos (Data Warehouse, DW)
  - Sistemas que almacenan datos históricos para ser utilizados por los Sistemas de Apoyo a la Decisión
  - Son sistemas de consulta enfocados a extraer conocimiento de los datos históricos almacenados
  - Basados en el análisis de los datos → On-Line Analytical Processing (OLAP)
    - Utilizan el modelado multidimensional (cubos, hipercubos, etc)



## El almacén de datos: primera aproximación

---

- Definición según W. Inmon (uno de los “pioneros”) (1992)
  - “Una colección de datos **orientados por tema, integrados, variables en el tiempo y no volátiles** que se emplea como apoyo a la toma de decisiones estratégicas”



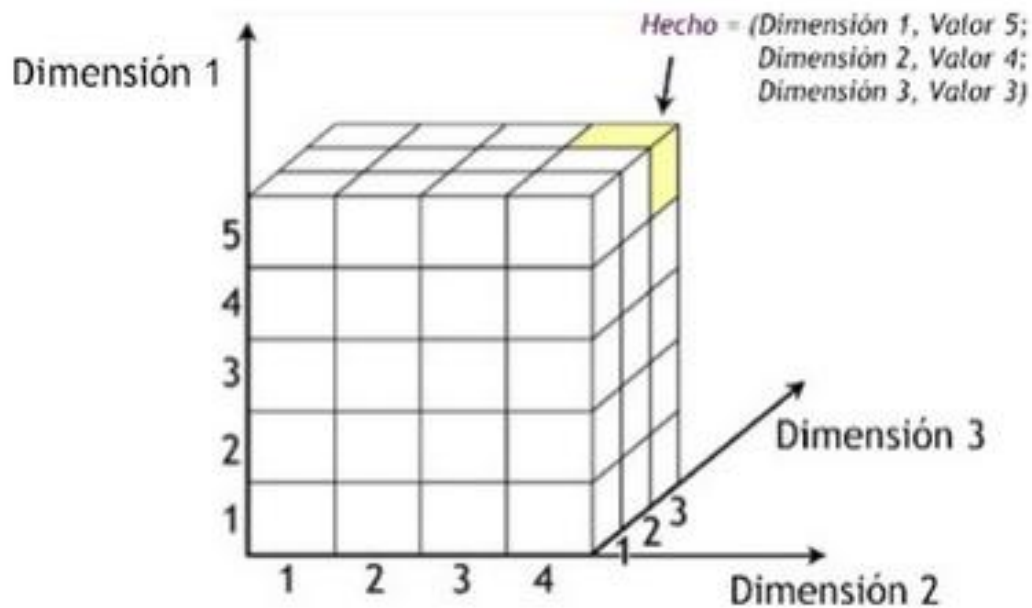


## El almacén de datos: primera aproximación

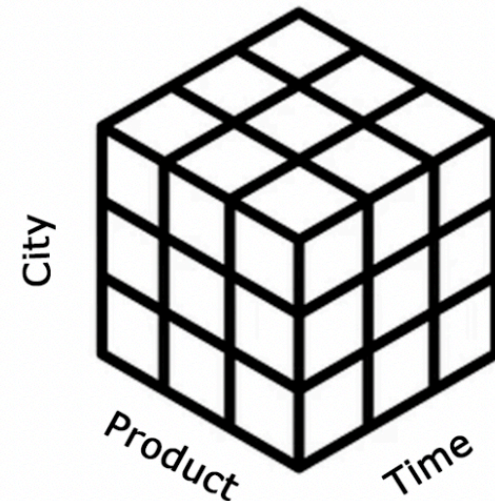
---

- Orientados por tema
  - El diseño enfocado a responder eficientemente a consultas estratégicas
    - Actividades de interés: compra, ventas, alquileres,...
    - Contexto de análisis: clientes, vendedores, productos, etc...
  - El modelado **Multidimensional** (primera aproximación)
    - **Hechos** → actividades de interés
    - **Dimensiones** → contexto de análisis

## El almacén de datos: primera aproximación



**Cubo multidimensional.**



P.e Hecho -> Ventas



## El almacén de datos: primera aproximación

---

- Integrados
  - Datos integrados de distintas fuentes de datos operacionales
- Variables en el tiempo
  - Datos relativos a un periodo de tiempo y se incrementan periódicamente
- No volátiles
  - Los datos almacenados no se modifican ni actualizan nunca, sólo se añaden nuevos datos



## El almacén de datos: primera aproximación

---

- Objetivo técnicas de almacenes de datos
  - Ofrecer información histórica para la toma de decisiones integrando la información procedente de distintas fuentes de datos operacionales (OLTP)



## El almacén de datos: primera aproximación

---

- Las tecnologías de almacenes de datos
  - Integran eficientemente tecnología de bases de datos con análisis de datos
    - Bases de Datos: Un SGBD que soporte el almacén
    - Análisis de datos: herramientas que permitan llevar a cabo un análisis sencillo de los datos
      - Las más extendidas: herramientas OLAP
      - Análisis Multidimensional basado en el modelo multidimensional



## El almacén de datos: primera aproximación

---

- Ventajas para las empresas
  - Decisiones soportadas por datos fiables
  - Rentabilidad de inversiones
  - Aumentar la competitividad en los nuevos entornos hostiles
  - Entornos amigables → los directivos analizan por ellos mismos los datos
- Por fin se entienden con los ordenadores → lo hemos conseguido !!!



## El almacén de datos: primera aproximación

---

- Inconvenientes
  - No planificar el alto esfuerzo para lograr un buen diseño
  - Nunca está acabado → incremento continuo de requerimientos ad-hoc



# Indice

---

- Repaso de la tecnología de bases de datos
- Nuevas necesidades de gestión de datos
- El Almacén de datos: primera aproximación
- **Diferencias entre sistemas operacionales y sistemas de almacenes de datos**





# Indice

---

- Repaso de la tecnología de bases de datos
- Nuevas necesidades de gestión de datos
- El Almacén de datos: primera aproximación
- **Diferencias entre sistemas operacionales y sistemas de almacenes de datos**

## Diferencias:

# Sistemas operacionales vs. Almacenes de datos

	<b>OLTP</b>	<b>DW/OLAP</b>
<b>Usuario</b>	▪ Profesional de TI	Analista de Información
<b>Función</b>	▪ Operaciones diarias	Apoyo a la decisión
<b>Diseño de BD</b>	▪ Orientada a la aplicación (Basado en EE-R)	Orientado al tema/negocio (estrella, Copos de nieve,...)
<b>Datos</b>	▪ Actuales, Aislados	Históricos, Consolidados
<b>Vistas</b>	▪ Detallados, Planos, Relac.	Agregados, Multidimensional
<b>Destino/utilización</b>	▪ Estructuradas, repetitivas	Ad-Hoc
<b>Unidades de trabajo</b>	▪ Transacciones simples	Consultas complejas
<b>Acceso</b>	▪ Lectura/escritura	Lectura mayoritariamente
<b># Registros accedidos</b>	▪ Decenas	Millones
<b># Usuarios</b>	▪ "Miles"	"Centenares"
<b>Tamaño de la BD</b>	▪ 100 MB-GB	100 GB-TB
<b>Medidas de rendimiento</b>	▪ Cantidad de transacciones	Cantidad de consultas, Respuesta



# El almacén de datos

## Bibliografía

---

- Juan Trujillo, Jose Norberto Mazón, Jesús Pardillo. [Diseño y explotación de almacenes de datos: Conceptos Básicos de Modelado Multidimensional](#). Ed. Club Universitario. 2011.
  - Tema 1
- Giovinnazo (2000). Object-Oriented Data Warehouse Design: Building a star schema
  - Tema 1 (introducción)
- Inmon (2002). Building the Data Warehouse (3ª ed.)
  - Tema 1 (introducción)
- Kimball (2002). The Data Warehouse Toolkit (3ª ed.)
  - Tema 1 (introducción)
- Thomsen (2000). OLAP solutions: Building Multidimensional Information Systems
  - Tema 1 (introducción)