

Federico Plancarte Sánchez

# E-data. Transformando datos en información con Data Warehousing

---

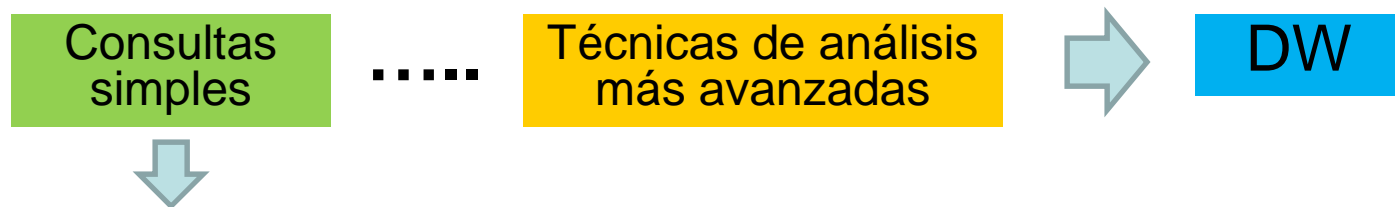
Tema 2

## El soporte a la Decisión



# Evolución del soporte a la decisión

## Diversas categorías del análisis del DS



Técnica de análisis más difundida (75%)

- Pasar de consultas simples a análisis avanzado
- ✓ Consulta simple: datos frescos, decisiones estratégicas
- ✓ Consultas “enlatadas” o predefinidas
- ✓ Consultas sin preparación (ad-hoc)

# ANALISIS MULTIDIMENSIONAL: El poder de dividir

---

Consulta estandar ➡ análisis multidimensional

- Técnica de análisis más poderosa
  - Hurgar más profundamente
- Diferentes perspectivas de los datos mediante el uso de “*dimensiones*”
- Información por una cierta *dimensión*
- *Drill down*
- Herramientas similares (a la consulta estándar)
- Mismos datos: diferentes formas

# Modelado y Segmentación: Análisis para trabajadores del conocimiento

---

Datos voluminosos  
y detallados



Trabajadores del  
conocimiento

*Modelo.* Colección de patrones para una característica dada

*Algoritmos* de modelado → Soporte a la decisión

La *segmentación* tampoco es una herramienta NUEVA

Ejemplo: Segmentación de CLIENTES

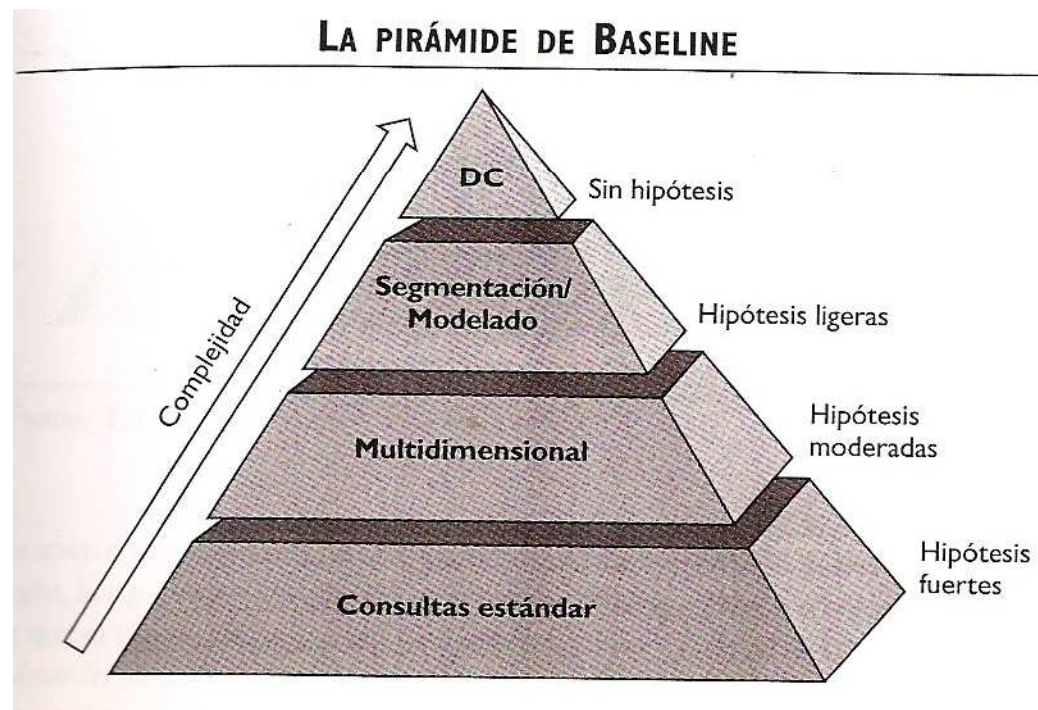
Software especializado

# Descubrimiento del Conocimiento

- *Marketing* → buscando los mejores clientes
- *KD*. Algoritmos poderosos que buscan patrones en grandes DB
- Patrones *no especificados* de antemano (como en el modelado)
- *DW* le dice a la empresa dónde están los *patrones y relaciones* importantes
- “*Respuestas desconocidas*” para encontrar nuevas formas innovadoras

# Algunas observaciones

- Evolución de la infraestructura del *DW*





# Algunas observaciones

---

- Evolución de la infraestructura del *DW*
- Los datos en DW evolucionan y maduran
- DW como una necesidad básica
  - Mayor frecuencia de uso
  - Ahorros de tiempo o costos
  - Consolidación de procesos

# Consultas Estándares

*“Listar los # de préstamos hechos al cliente X y las fechas de pago cuando se ha demorado más de una semana”*

La consulta busca *confirmar* un supuesto  
⇒ parte de una *hipótesis muy fuerte*

*“Listar a todos los clientes para los que el ingreso por uso a la hora pico haya disminuido un 20% o más”*

Resultado ⇒ acciones comerciales





# Consultas Estándares

---

*“Mostrar a todos los clientes de celulares con llamadas entrantes incompletas en más de un 20% en cada semana”*

*Resultados útiles:*

# Análisis multidimensional

*“Mostrar a todos los clientes de celulares con llamadas entrantes incompletas en más de un 20% en c/semana”*

## *Drill-down*

*“De esos clientes ¿quiénes usan roaming por fuera de la red?”*

Resultados:



# Análisis multidimensional

*“Mostrar los ingresos trimestrales correspondientes a grandes clientes comerciales en las regiones norte, NO y SO en 1977 y 1978”*

## UN INFORME BÁSICO PARA SOPORTE DE DECISIÓN

	1997				1998	
	1 <sup>er</sup> trim.	2 <sup>do</sup> trim.	3 <sup>er</sup> trim.	4 <sup>to</sup> trim.	1 <sup>er</sup> trim.	2 <sup>do</sup> trim.
Norte	50	35	44	95	45	33
Noroeste	38	43	43	50	38	45
Sudoeste	34	23	46	72	30	28

Ingresos contabilizados (en miles)

# Análisis multidimensional

*“Mostrar los mismos datos por distrito dentro de la región noroeste”*

## SALIDA DEL OLAP

		1997				1998	
		1 <sup>er</sup> trim.	2 <sup>do</sup> trim.	3 <sup>er</sup> trim.	4 <sup>to</sup> trim.	1 <sup>er</sup> trim.	2 <sup>do</sup> trim.
Norte		50	35	44	95	45	33
Noroeste	Distrito A	13	15	14	18	11	14
	Distrito B	10	12	12	13	8	16
	Distrito C	15	16	17	19	19	15
Sudoeste		34	23	46	72	30	28

Ingresos contabilizados (en miles)

# Análisis multidimensional

*“Mostrar los mismos datos para el distrito B de la región noroeste”*

## ANÁLISIS EN PROFUNDIDAD (DRILL-DOWN)

		1997			1998		
		1 <sup>er</sup> . trim.	2 <sup>do</sup> . trim.	3 <sup>er</sup> . trim.	4 <sup>to</sup> . trim.	1 <sup>er</sup> . trim.	2 <sup>do</sup> . trim.
Norte		50	35	44	95	45	33
Noroeste	Distrito B	10	12	12	13	8	16
	Toledo	6	7	6	6	4	9
	Columbus	2	2	2	3	2	3
	Cleveland	2	2	2	3	2	3
Sudoeste		34	23	46	72	30	28

Ingresos contabilizados (en miles)



# Análisis multidimensional

---

*Las herramientas OLAP son únicas para facilitar al usuario pedir los mismos datos en diferentes formas*

# Análisis multidimensional

## *PROVEEDORES*

- *Oracle OLAP*
- *SAP - OLAP*
- *Microsoft*
- *Inesoft*
- *Dundas Chart for .NET OLAP Services*
- *MicroStrategy OLAP Services*

# Modelado y Segmentación

## Modelado

- *Comportamiento futuro de los clientes*
- *Su viabilidad a largo plazo*

## Segmentación

- *Clasificar y reclasificar clientes por*
  - *Demografía*
  - *Patrones de compra*
  - *Proclividad a la compra*





# Modelado y Segmentación

## Ejemplos de Modelado

- *Valor de por vida del cliente*
- *Desgaste del cliente*
- *Modelado predictivo*



# Modelado y Segmentación

## Ejemplos de Segmentos

- *Clientes que responden a ofertas*
- *Clientes que responden a descuentos*
- *Clientes que responden a productos nuevos*
- *Clientes que responden a promociones*



# Modelado y Segmentación

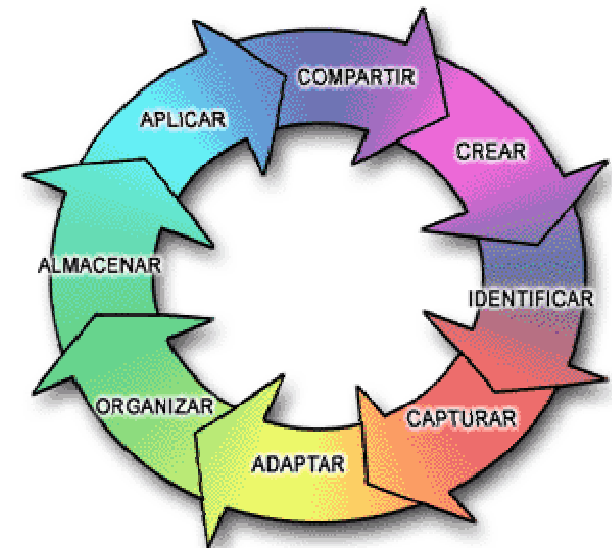
## Resultados de la segmentación:

- *¿A qué grupo de clientes enfocar un nuevo servicio?*
- *¿Quién es más probable que se interese por este servicio?*
- *¿Qué clientes es más probable que cometan fraude?*
- *¿Qué clientes es más probable que respondan a descuentos?*



# Descubrimiento del conocimiento

- *Cero hipótesis*
- *Encuentra en los datos patrones ocultos:*
  - *Comportamiento del cliente*
  - *Ventas por producto*
  - *Cancelaciones*
  - *Compras futuras*



## *Ejemplos:*

- *Qué pasa cuando 2 medicamentos se toman juntos*
  - *Patrones desconocidos de afinidad de productos*
  - *Productos disparadores de otros*

# Descubrimiento del conocimiento

*Más ejemplos:*

*De los clientes que compran frituras, el 66% también compran golosinas*

*De los clientes que compran frituras y golosinas, el 75% también compra vino tinto*



## ACCIONES

- *Enviar cupones a compradores frecuentes*
- *Eliminar descuentos en vino tinto*
- *Reubicar los productos dentro de la tienda*



# Descubrimiento del conocimiento

---

*Ejemplos de datos que salen a la luz gracias a KD:*

- *Ciertos productos disparadores afectan otras compras*
- *Descubrir la “próxima compra probable”*
- *Descubrir patrones en la caída de compras*
- *Aspectos del cliente que afectan el ciclo de vida del producto*



# Data mining

---

- *El Data mining comprende la segmentación, el modelado y el descubrimiento del conocimiento*
- *Es sinónimo de análisis estadístico*
- *Está soportado por 3 tecnologías*
  - Recolección masiva de datos
  - Potentes computadoras con multiprocesadores
  - Algoritmos de Data Mining



# Data mining

---

Se desarrolla bajo *lenguajes de última generación* basados en *inteligencia artificial* y utiliza *modelos matemáticos* tales como:

- Redes neuronales artificiales
- Árboles de decisión
- Reglas de inducción
- Algoritmos genéticos





# Data mining

---

Predice *futuras tendencias y comportamientos* para permitir en los negocios tomar decisiones proactivas

Estas herramientas pueden responder a preguntas de negocios que *consumen demasiado tiempo*

Exploran las bases de datos en busca de *patrones ocultos*, encontrando información predecible que un experto *no puede llegar a encontrar* porque está fuera de sus expectativas



# Data Warehouse en la realidad

---

Los datos residen en la menor cantidad de *plataformas distintas*, preferentemente en un *DW* o *Data mart*

Los gerentes acostumbran *acceder directamente al DW* cuando requieren información

Los ejecutivos apoyan el *soporte a la decisión*

El *DW* puede *mejorar la productividad y aumentar la rentabilidad* de la organización



# Referencias

---

E-data. Transformando los datos en información con Data Warehousing  
Capítulo 2. El soporte a la Decisión de abajo hacia arriba  
Jill Dyché