

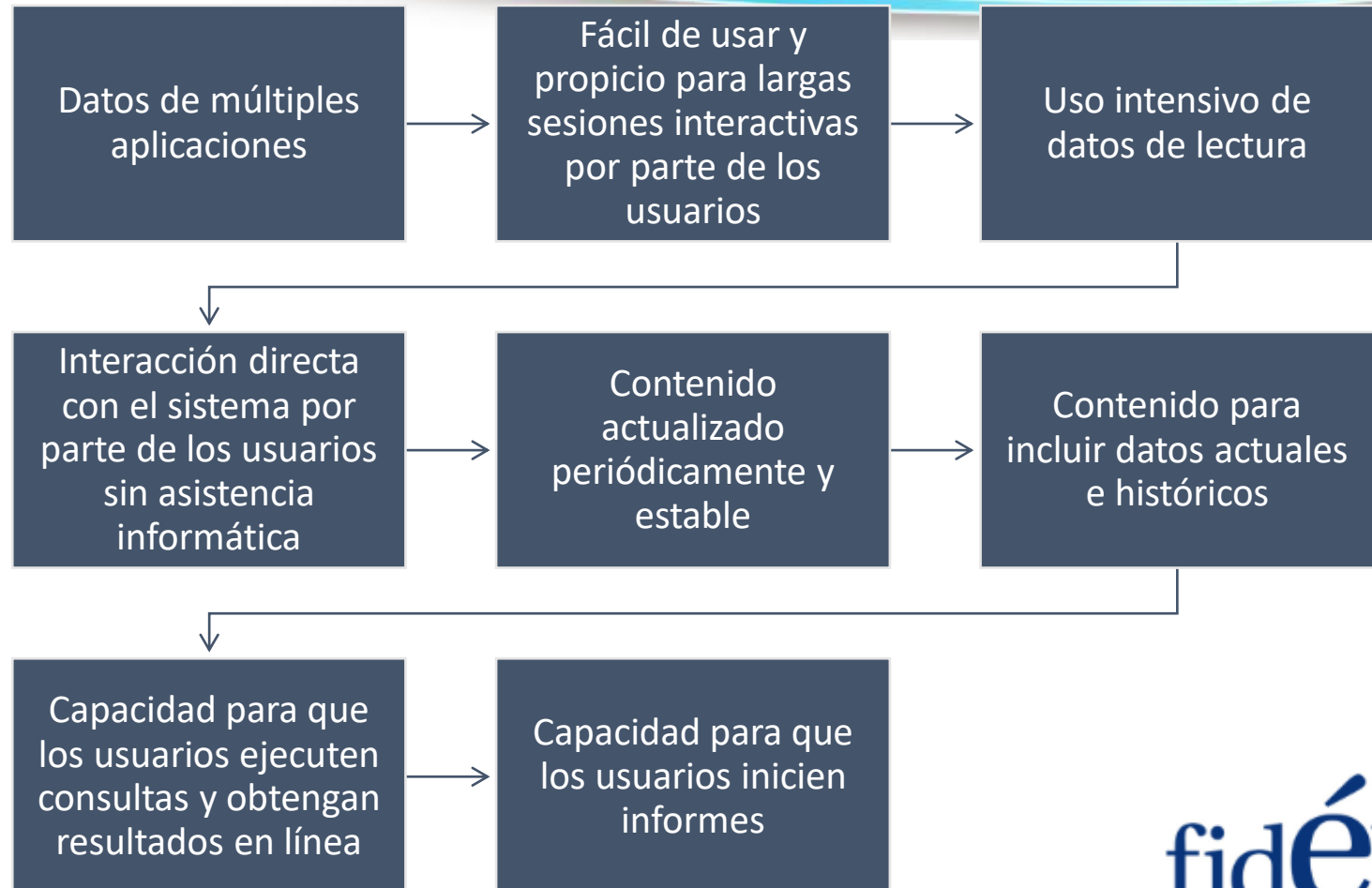
# SC-602 Data Warehouse y bases de datos

Lección número 3  
Tendencias en Data Warehouse

# Agenda

- Repaso de la clase anterior y dudas
- Tendencias en Data Warehouse
  - Crecimiento continuo de Data Warehouse
  - Tendencias significativas
  - Estándares emergentes
  - Web-Enabled Data Warehouse
- Contactos administrativos

## Funciones que debe cumplir DW



# Niveles requeridos de análisis

- Ejecución de consultas e informes simples contra datos actuales e históricos.
- La capacidad de realizar análisis "qué pasaría si" es de muchas maneras diferentes
- Habilidad para consultar, retroceder, analizar y luego continuar el proceso a la longitud deseada
- Detectar tendencias históricas y aplicarlas para resultados futuros.



## BI y DW

### OPERATIONAL SYSTEMS



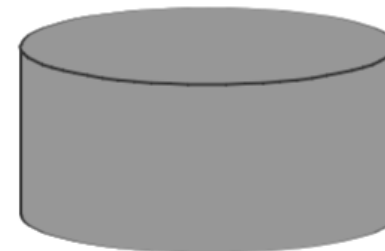
Basic  
business  
processes

Extraction,  
cleansing,  
aggregation

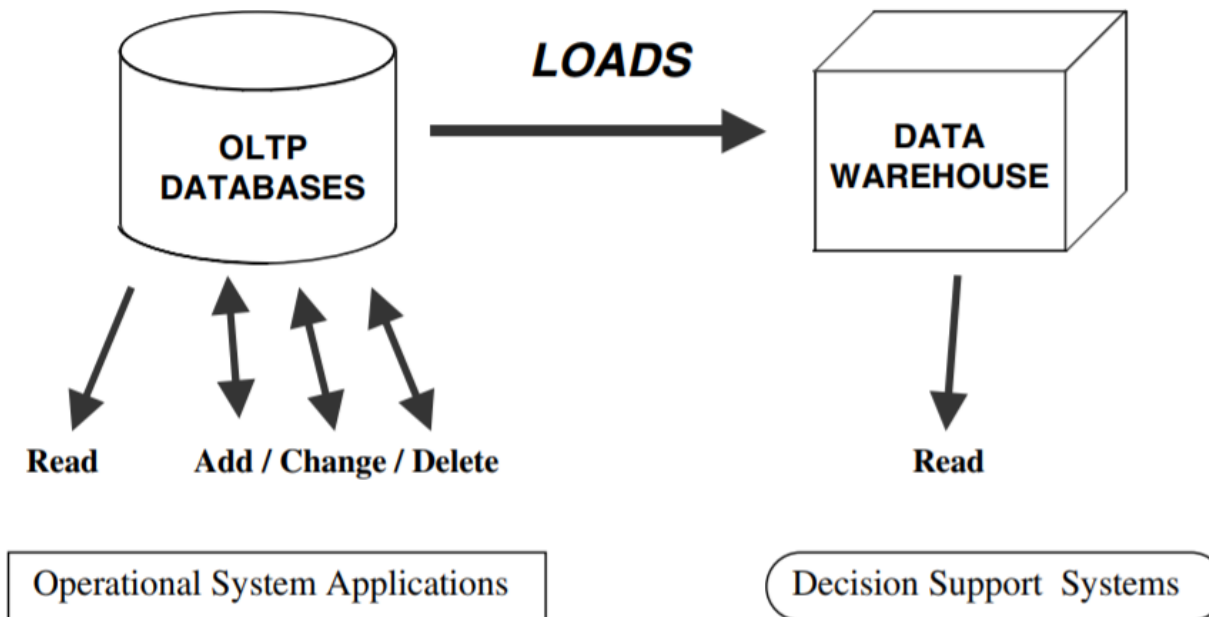


**Data Transformation**

Key measurements,  
business dimensions



Usually the data in the data warehouse is not updated or deleted.



# Crecimiento continuo del DW

- Competición feroz
- Desregulación del gobierno
- Necesidad de modernizar los procesos internos.
- Imperativo para la comercialización personalizada.

**Sistemas bancarios y telecomunicaciones llevan la delantera.**

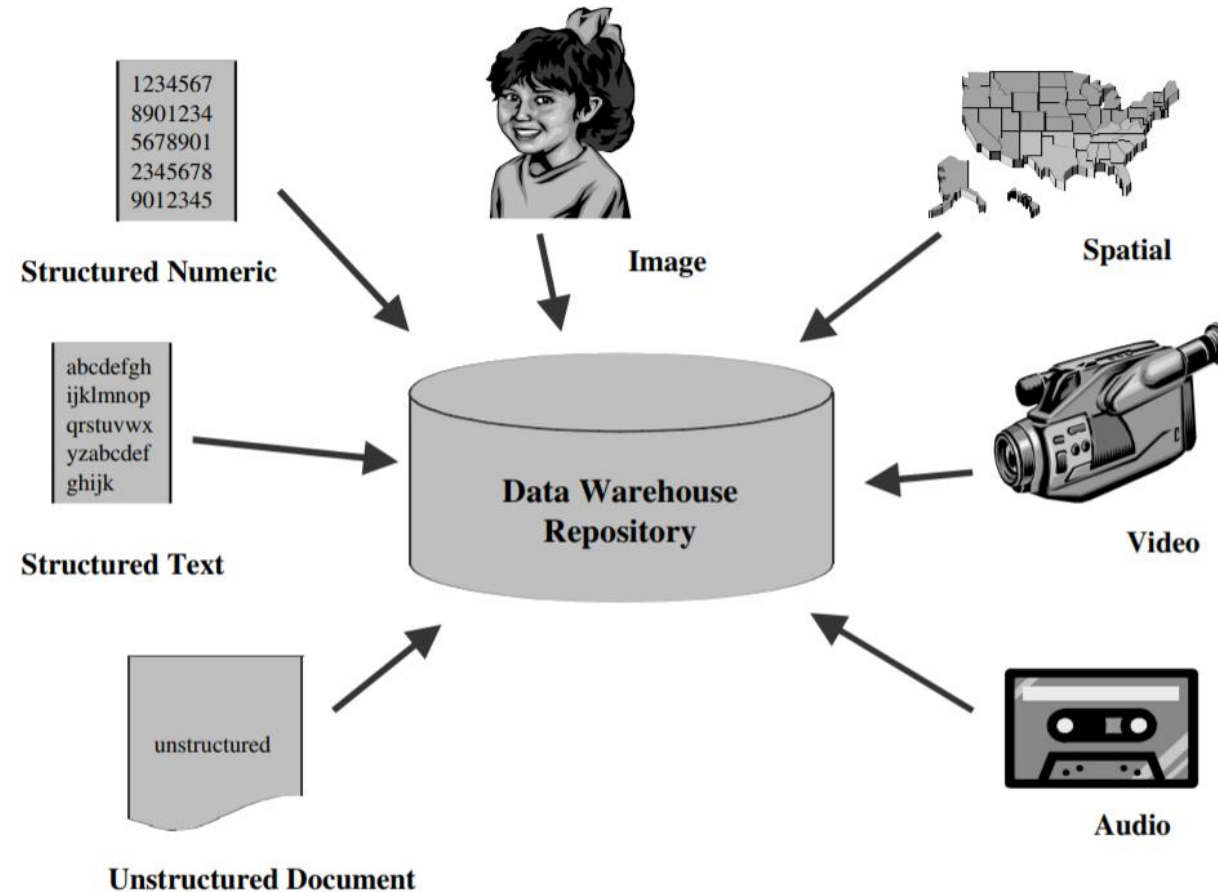


# Tendencias significativas





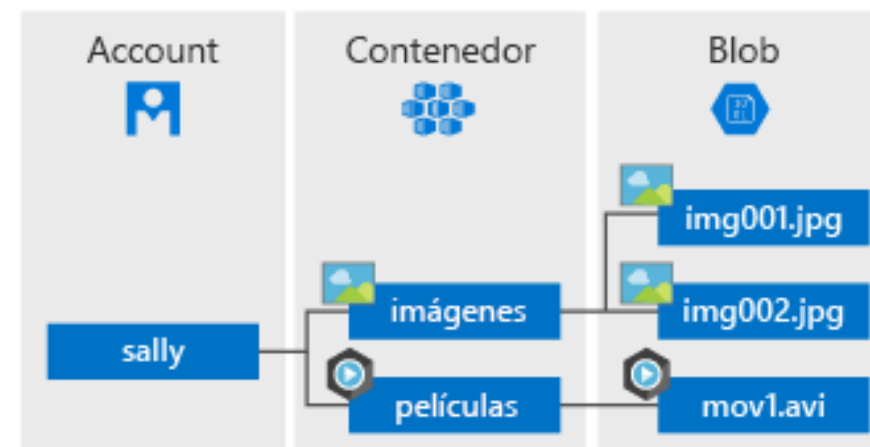
## Data types





## Datos no estructurados

- Mayormente videos e imágenes almacenados como objetos binarios grandes (BLOBs).
- No siempre pueden almacenarse como un tipo de datos relacional.
- Por ejemplo, un videoclip requeriría un servidor que admita la entrega de múltiples transmisiones de video a una dada velocidad y sincronización con la parte de audio.



# ¿Cómo consultarlos?



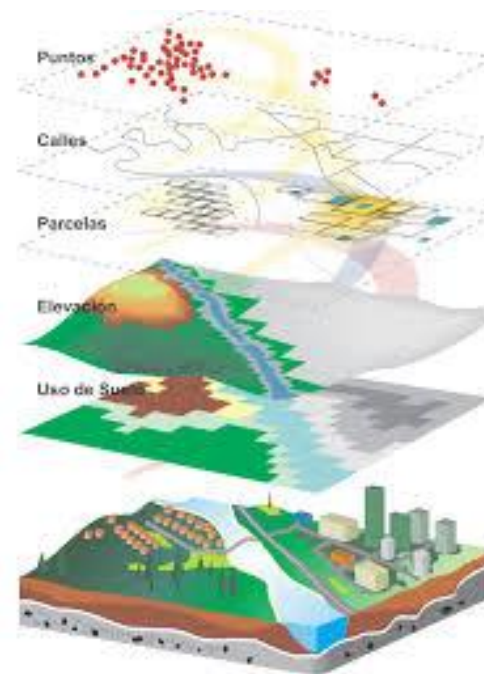
- Nuevos motores de búsqueda especializados en datos no estructurados.
- La consulta por contenido de imagen es un ejemplo de un mecanismo de búsqueda de imágenes. El producto le permite preindexar imágenes basadas en formas, colores y texturas.
- Búsquedas por palabras, cadenas de caracteres, frases, comodines, operadores de proximidad y booleanos





## Datos espaciales

- Estas bases de datos utilizan un cuadro de referencia como el Sistema de Referencia Espacial para guardar datos como dirección, calle, cuadrante de la ciudad, condado, estado y zona.





# Visualización de datos

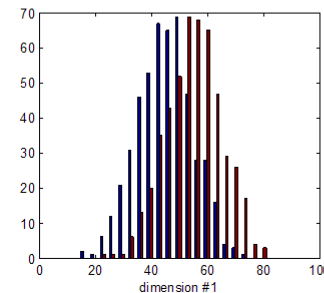
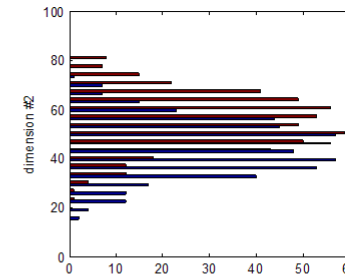
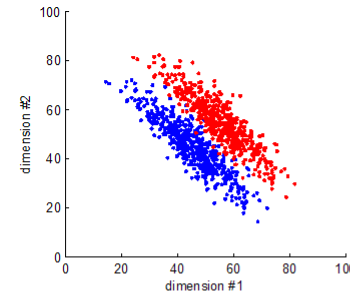
- Más tipos de gráficos
- Visualización interactiva
- Visualización de conjuntos de resultados complejos y grandes
- Manipulación de gráficos
- Mayor nivel de detalle
- Interacción avanzada





## Otros...

- Procesamiento paralelo
- Herramientas para *Queries*
- Herramientas de navegación
- Fusión de datos





# DW integrado

Los ***Enterprise Resource Planning*** abarcan un mercado de más de \$45 billones. Soportan desde inventario hasta facturación, desde recursos humanos hasta producción.

Los sistemas de ***Knowledge Management*** gestionan eficientemente la información pero suele guardarse en datos no estructurados.

Finalmente, cada vez es más importante la retención de clientes por medio del ***Customer Relationship Management*** y combinando el conocimiento de cada contacto que se ha tenido con un cliente.





**ERP**

**CRM**

**Financial**

**Knowledge  
Management**

**SoftClouds  
ERP  
Expert**

**SoftClouds  
CRM  
Expert**

**SoftClouds  
Financial  
Expert**

**SoftClouds  
KM  
Expert**

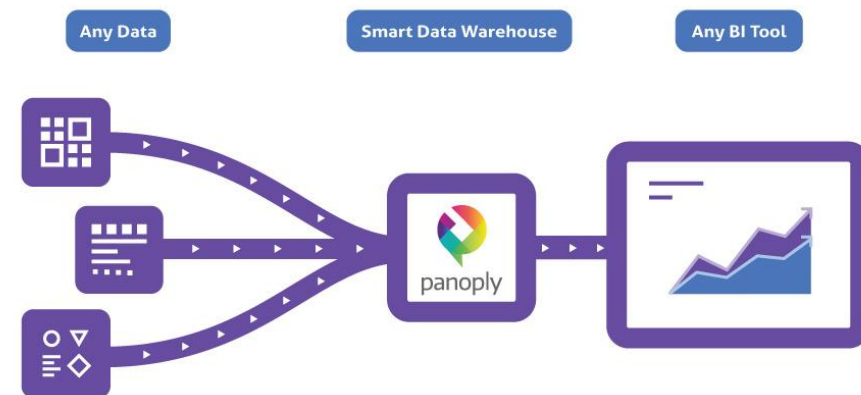




## Web Enabled DW

Las empresas no pueden sobrevivir sin presencia web.

La combinación de un uso inteligente de los datos y las herramientas que proporciona la web puede ser muy poderosa.





## Del Data Warehouse a la web

Anteriormente se mantenía una estructura cliente-servidor para usuarios específicos de toma de decisiones (usuarios internos).

Para incrementar la productividad de toda la cadena de valor, se necesita incluir a clientes, proveedores y otros pares de negocio. Incluso, al público en general.

La web se convierte en el mecanismo principal de tratado de información.

Datos Self-Service

Análisis Interactivo

Alta disponibilidad

Alto rendimiento

# De la web al Data Warehouse

Captar el *clickstream* y aprovecharlo.

- Demanda del cliente
- Eficacia de las promociones de marketing.
- Eficacia de la relación de afiliación entre productos.
- Recopilación de datos demográficos.
- Patrones de compra



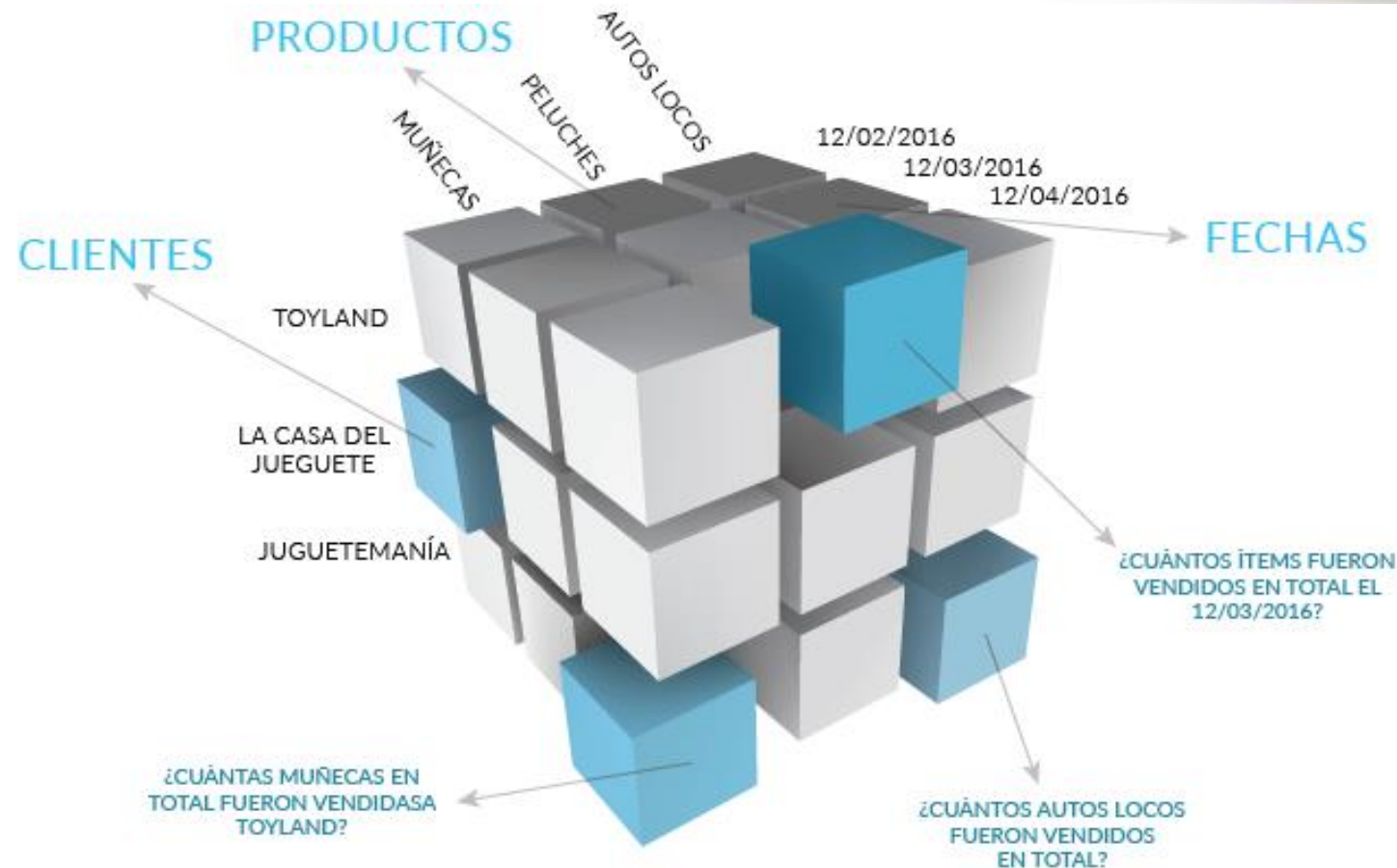


# Modelos



# Modelos arquitectónicos

- **Enterprise Warehouse** (único DW para toda la organización).
- **Data Marts** (varios DW para grupos específicos de usuarios).
- **Virtual Warehouse** (vistas sobre las bases de datos operativas, de las cuales sólo se materializan algunas de ellas).



## Modelo multidimensional

# La multidimensión

Modelos de datos como conjuntos de medidas descritas por dimensiones.

- Adecuado para resumir y organizar datos (generalización de las hojas de cálculo).
- Enfocado para trabajar sobre datos de tipo numérico.
- Más simple, más fácil de visualizar y de entender que el modelado E/R.



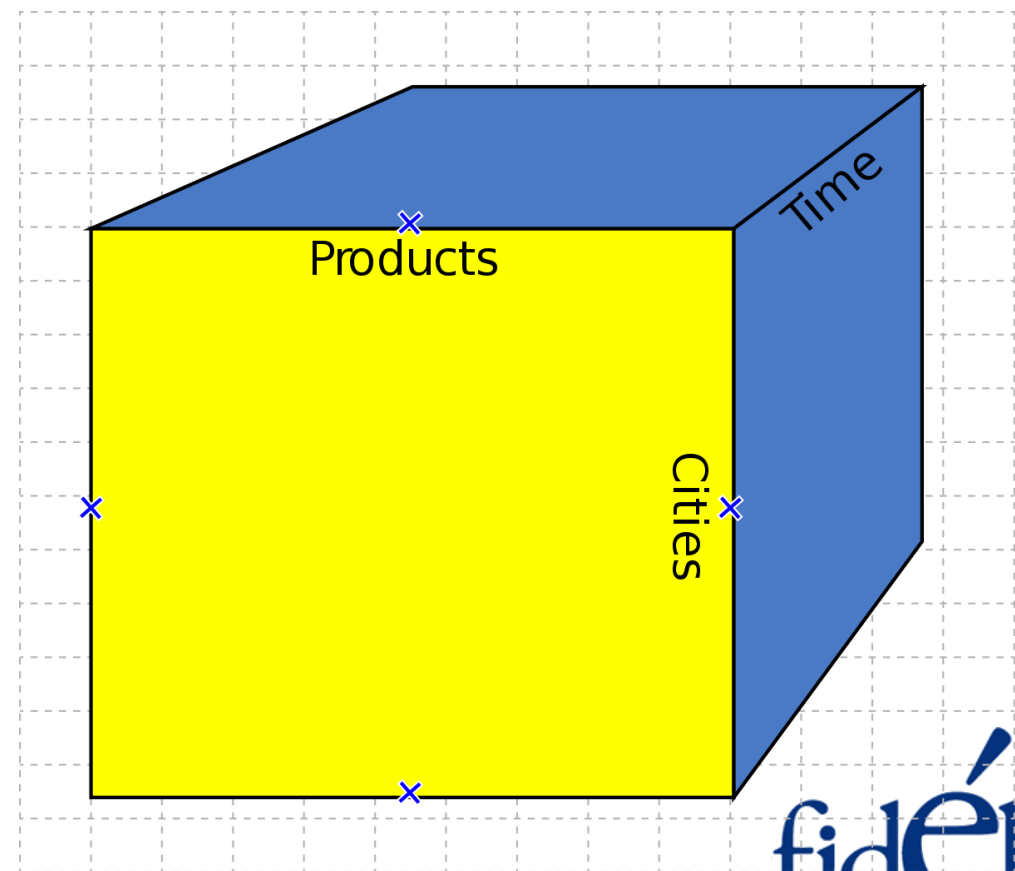


## Dimensiones

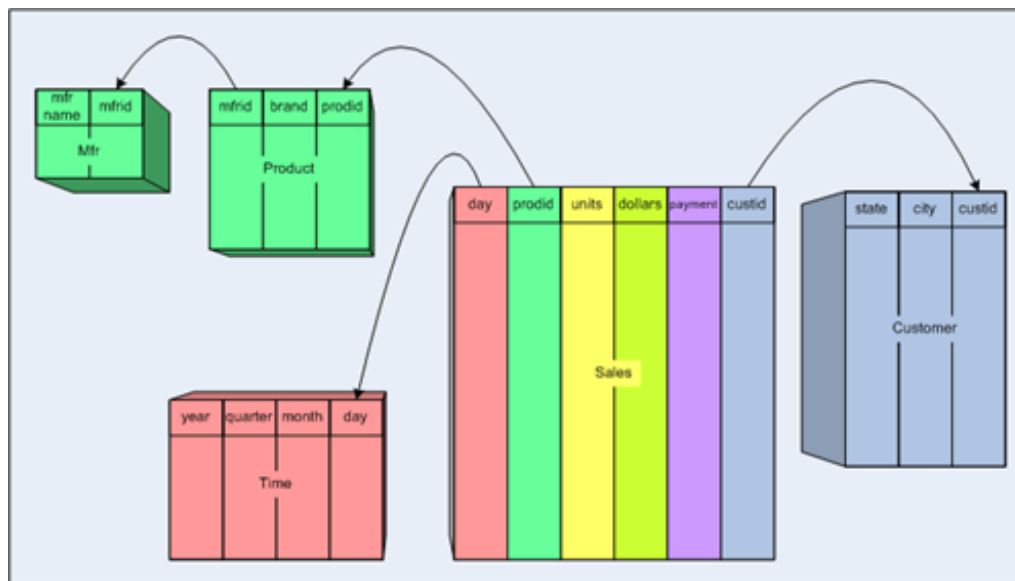
Perspectivas o entidades respecto a las cuales una organización quiere mantener sus datos organizados.

### Ejemplos:

- Tiempo
- Localización
- Clientes
- Proveedores
- Productos







## Miembros

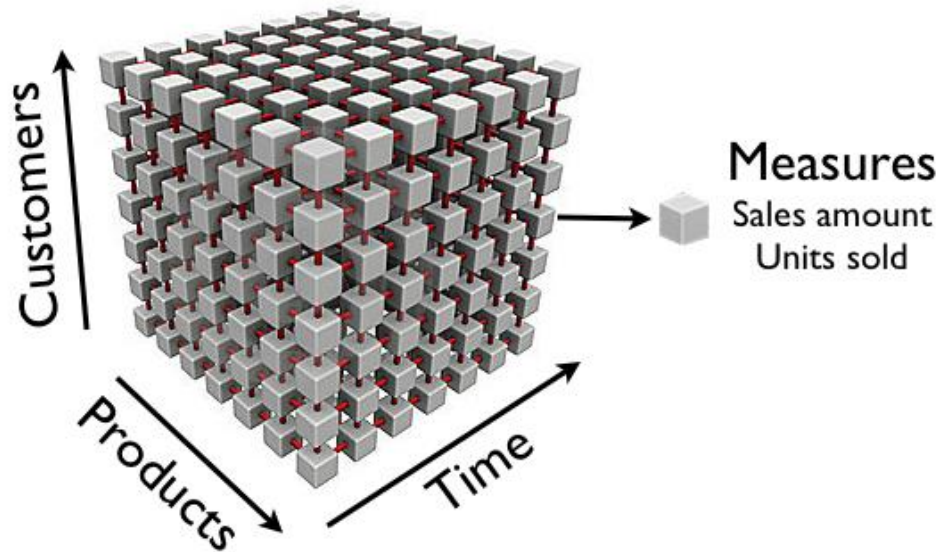
Nombres o identificadores que marcan una posición dentro de la dimensión.

### Ejemplos:

- Meses, trimestres y años son miembros de la dimensión **tiempo**.
- Ciudades, regiones y países son miembros de la dimensión **localización**.

## Medida

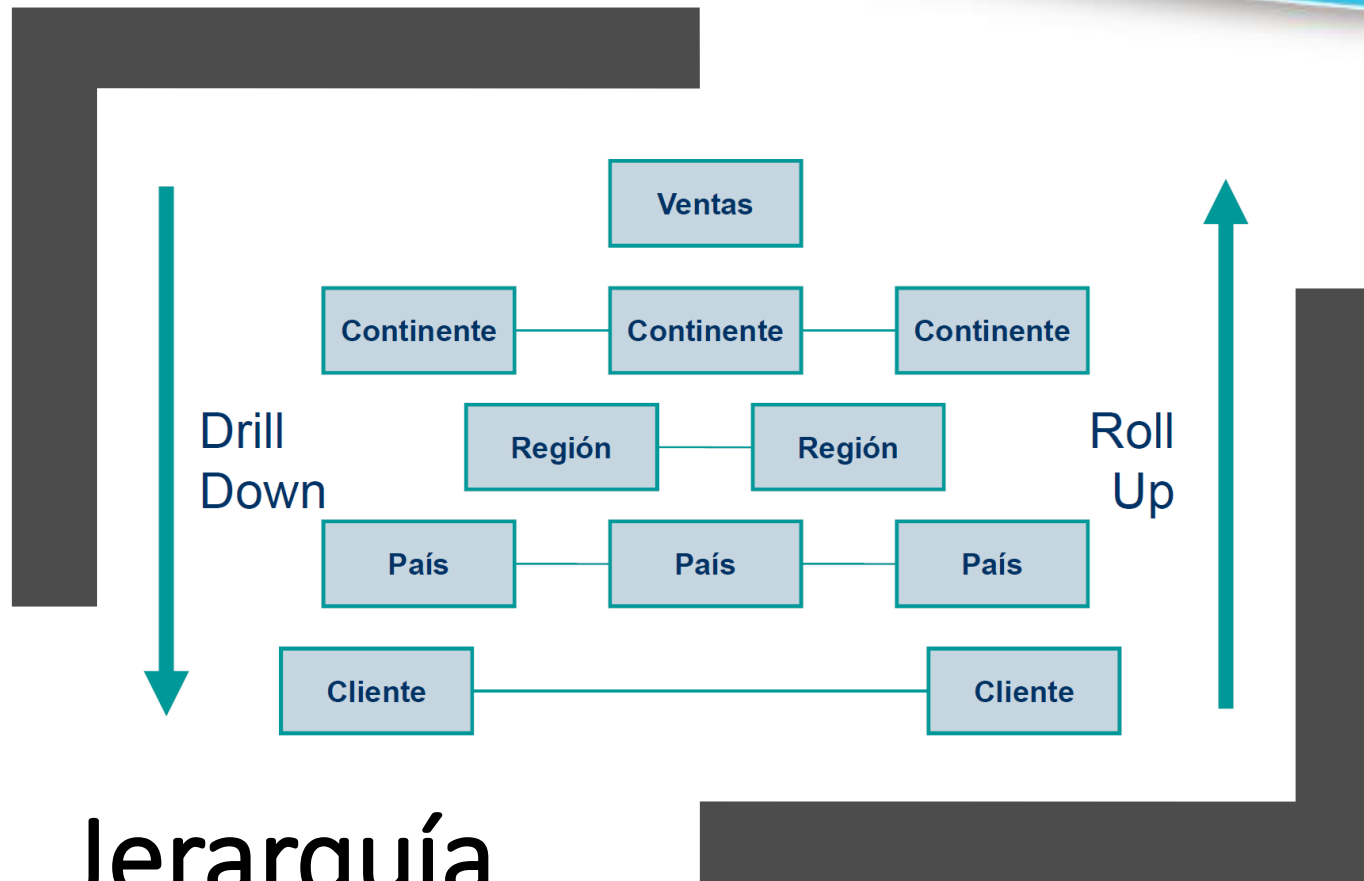
Atributos numéricos asociados a los hechos (lo que realmente se mide).



## Hechos

Colecciones de datos relacionados compuestas por medidas y un contexto.

- Las dimensiones determinan el contexto de los hechos.
- Cada hecho particular está asociado a un miembro de cada dimensión.



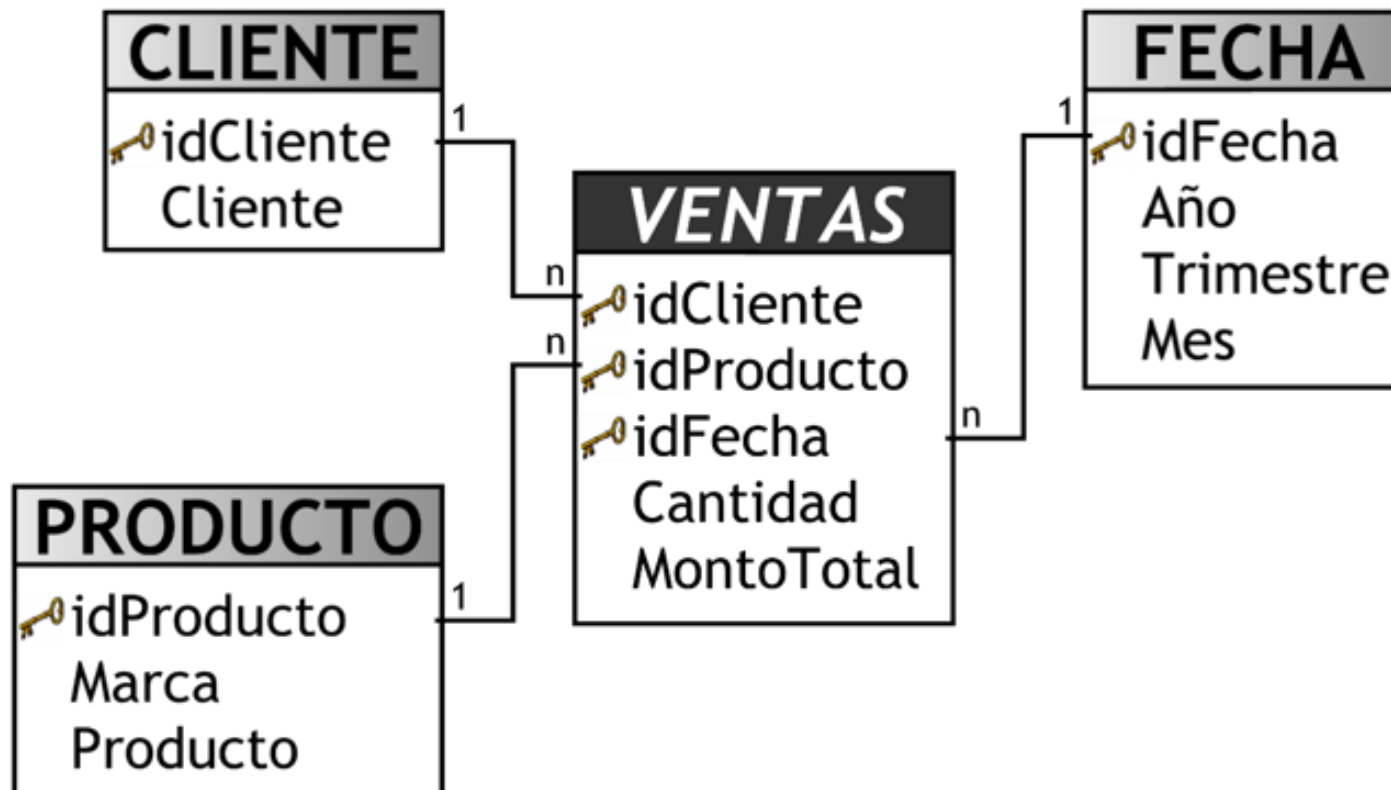
## Jerarquía

Los miembros de las distintas dimensiones se suelen organizar en forma de jerarquías.



## Ejemplo

Creación de un cubo de datos.



# Indicadores o medidas

- Unidades vendidas
- Monto total de ventas



## Cubo de Ventas

- Unidades Vendidas |  $SUM([VENTAS.Cantidad])$
- Monto Total de Ventas |  $SUM([VENTAS.MontoTotal])$

Tomamos los hechos Ventas Cantidad y Ventas MontoTotal



## Atributos



### Cubo de Ventas

- Clientes | [CLIENTE.Cliente]
- Marcas | [PRODUCTO.Marca]
- Productos | [PRODUCTO.Producto]
- Años | [FECHA.Año]
- Trimestres | [FECHA.Trimestre]
- Meses | [FECHA.Mes]
- Unidades Vendidas | SUM([VENTAS.Cantidad])
- Monto Total de Ventas | SUM([VENTAS.MontoTotal])

Tomamos de las dimensiones los campos necesarios.

Note que para la dimensión Fecha, podemos tomar varios miembros de su jerarquía: Años, Trimestres, Meses...

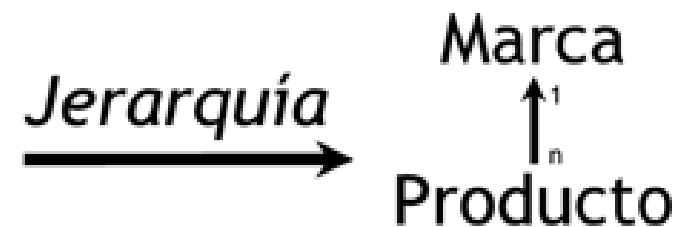


## Jerarquías

- Jerarquía Productos

Se aplicarán a los Atributos:

- Marcas
- Productos



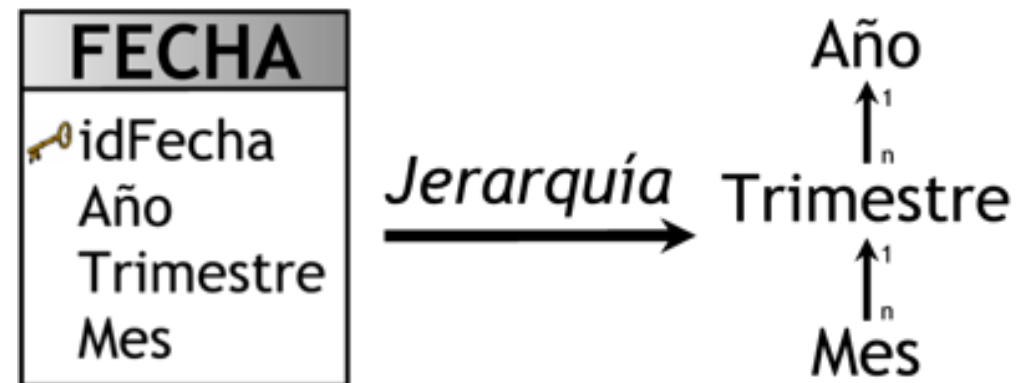
Un producto en especial pertenece solo a una marca. Una marca puede tener uno o más productos.

## Jerarquías

- Jerarquía Fechas

Se aplicarán a los Atributos:

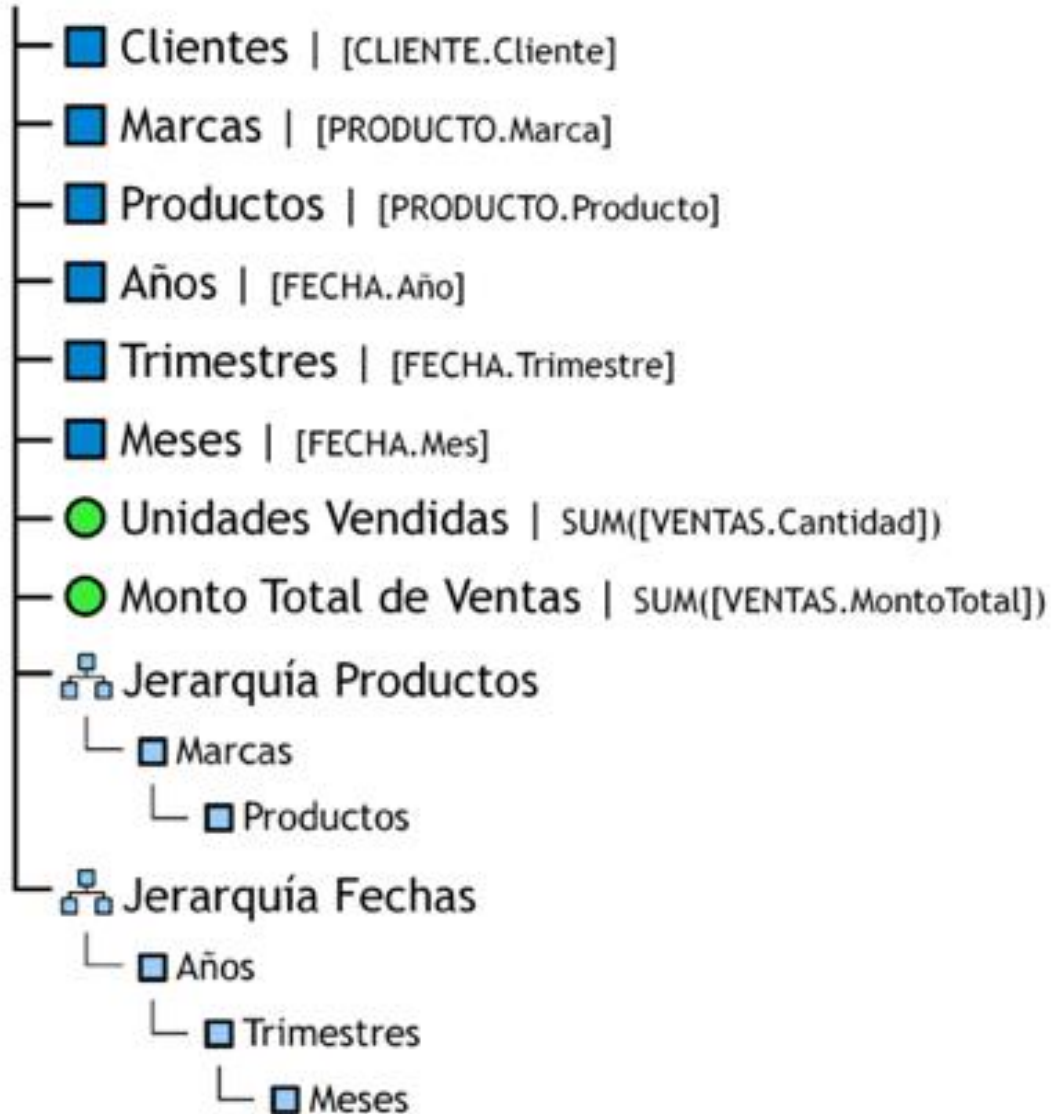
- Años
- Trimestres
- Meses



- Un mes del año pertenece solo a un trimestre del año. Un trimestre del año tiene uno o más meses del año.
- Un trimestre del año pertenece solo a un año. Un año tiene uno o más trimestres del año.



## Cubo de Ventas



# Cubo final





# Contactos administrativos

Profesor del curso: Miguel Corea ( [mcorea797@ufidelitas.ac.cr](mailto:mcorea797@ufidelitas.ac.cr) )

Subdirectora: Ing. Katya Brenes ( [kbrenes@ufidelitas.ac.cr](mailto:kbrenes@ufidelitas.ac.cr) )

Director: Ing. Wilberth Molina. MSc ( [wmolina@ufidelitas.ac.cr](mailto:wmolina@ufidelitas.ac.cr) )