

Teoría de Bases de Datos

NOCIONES BÁSICAS

JAIME MIRANDA

Departamento de Ingeniería Industrial
Universidad de Chile

Nacieron en los 70'S creadas por E.F.Codd.

Están basadas en relaciones (tablas) como estructura de almacenamiento, con atributos o campos (columnas) y una serie de tuplas o registros (filas).

Estandarizaron el lenguaje de manipulación, usando SQL, creado por IBM en los 80'S.

ALGUNOS EJEMPLOS...

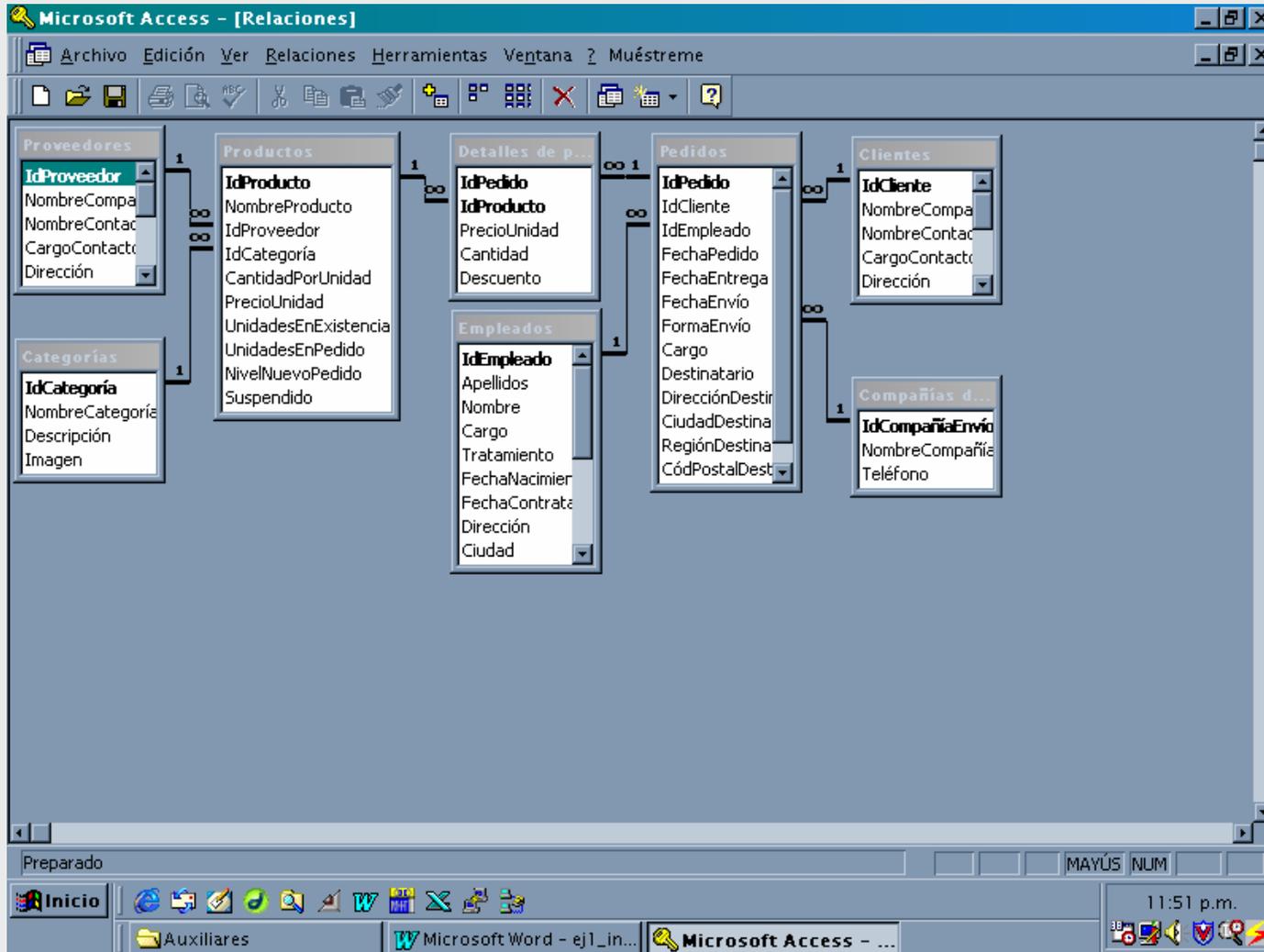
Relación Empleado

EMPLEADO	NPILA	APPAT	APMAT	<u>RUT</u>	FNAC	DIRECCION	SEXO	SUELDO	RUTSUPER V	NDEPTO
	Juan	Perez	Martinez	13.463.530-4	12-01-78	Av.Matta 223	M	120.000	123654	5
	Alicia	Rubio	Jara	15.356.345-8	25-06-65	Alameda 123	F	190.000	852647	4
	Sebastian	Carrasco	Claro	10.254.269-7	18-12-50	San Diego 654	M	250.000	843601	1

Relación Departamento

DEPARTAMENTO	DNOMBRE	<u>DNUMERO</u>	RUTGERENTE	GERFECHAINIC
	Of. Central	1	88866555	19-06-71
	Administración	4	98765432	01-01-85
	Investigación	5	33344555	22-05-78

ALGUNOS EJEMPLOS ...



SINTAXIS SQL : “Structured Query Language”

→ SELECT < Lista de atributos >

→ FROM < Lista de tablas >

→ WHERE < Condición >

Recuperar todos los números de RUT de los empleados.

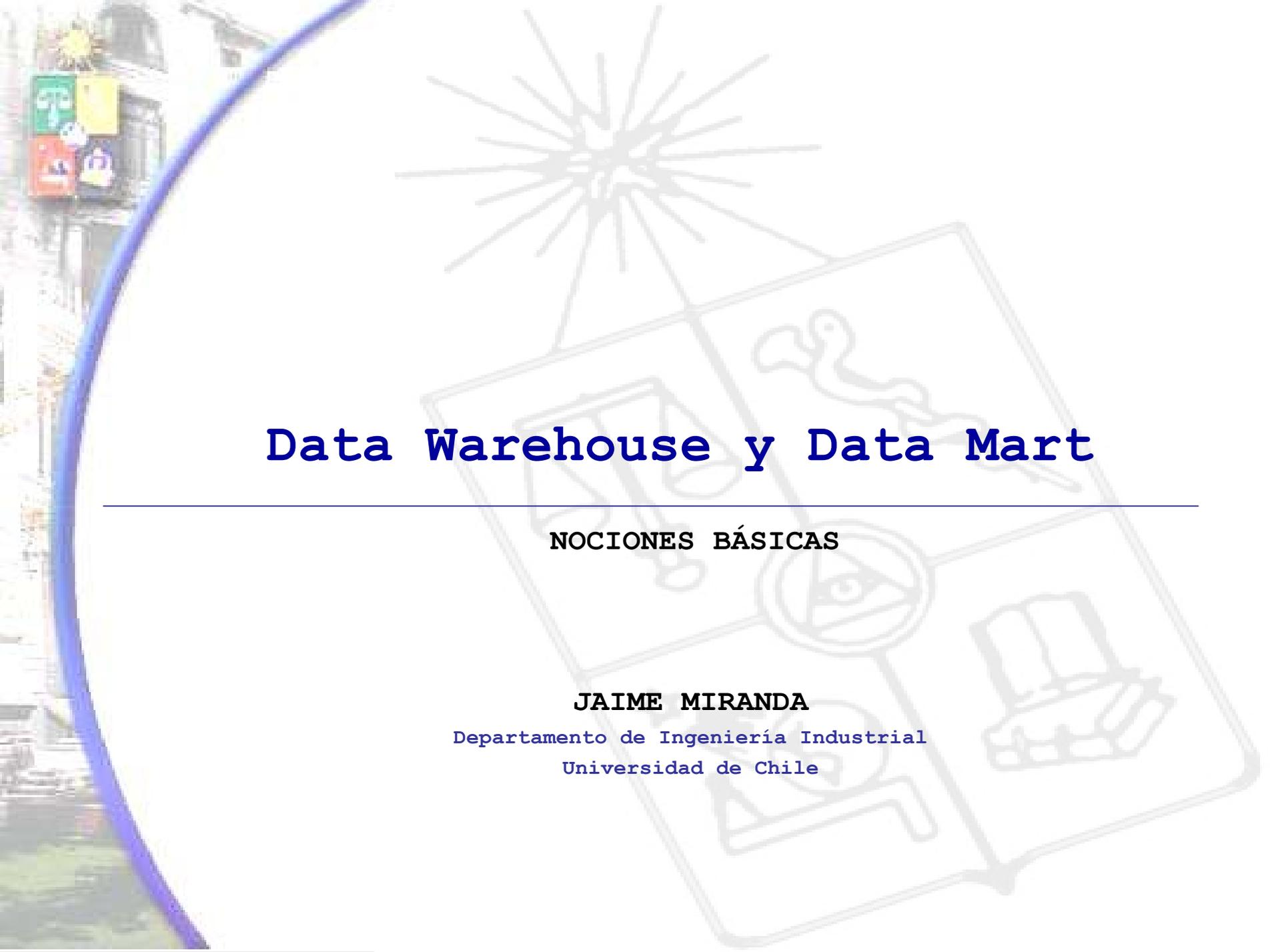
- SELECT RUT
- FROM EMPLEADO

Recuperar los valores de todos los atributos de EMPLEADO que trabajan en el departamento número "5".

- SELECT *
- FROM EMPLEADO
- WHERE NDEPTO = 5

Comando AND: Recuperar la fecha de nacimiento y dirección del empleado Juan Pérez

- SELECT FNAC, DIRECCION
- FROM EMPLEADO
- WHERE NPILA = "Juan" **AND** APPAT = "Pérez"



Data Warehouse y Data Mart

NOCIONES BÁSICAS

JAIME MIRANDA

Departamento de Ingeniería Industrial
Universidad de Chile

ALGUNAS FUNCIONES...

Reúne datos esenciales provenientes de bases de datos heterogéneas desde todas las áreas de negocio (Ventas, finanzas, RRHH, etc.)

Organiza los datos para apoyar decisiones de gestión.

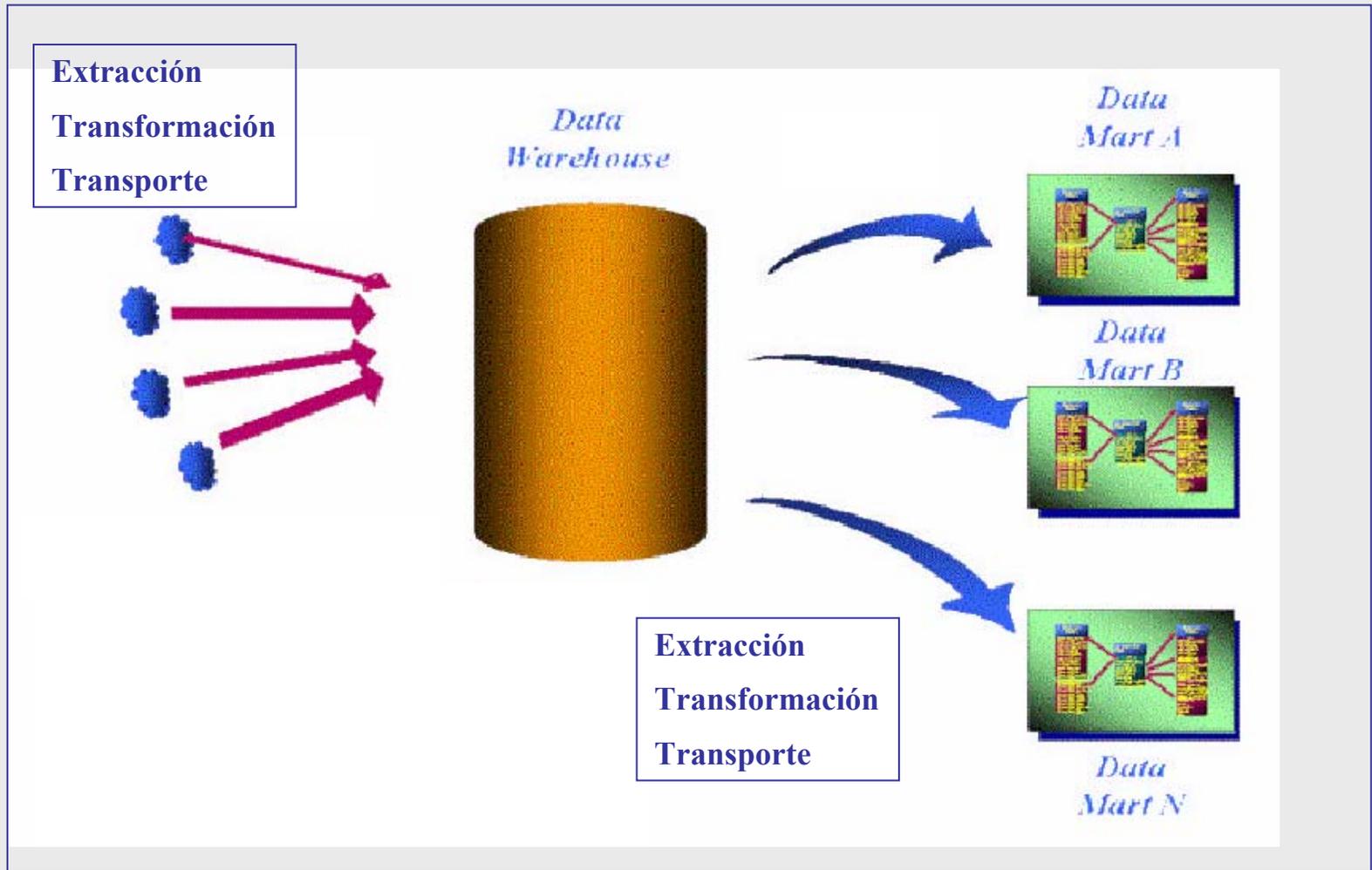
Maneja elevados volúmenes de información.

Permite el mejor funcionamiento de los métodos de Data Mining.

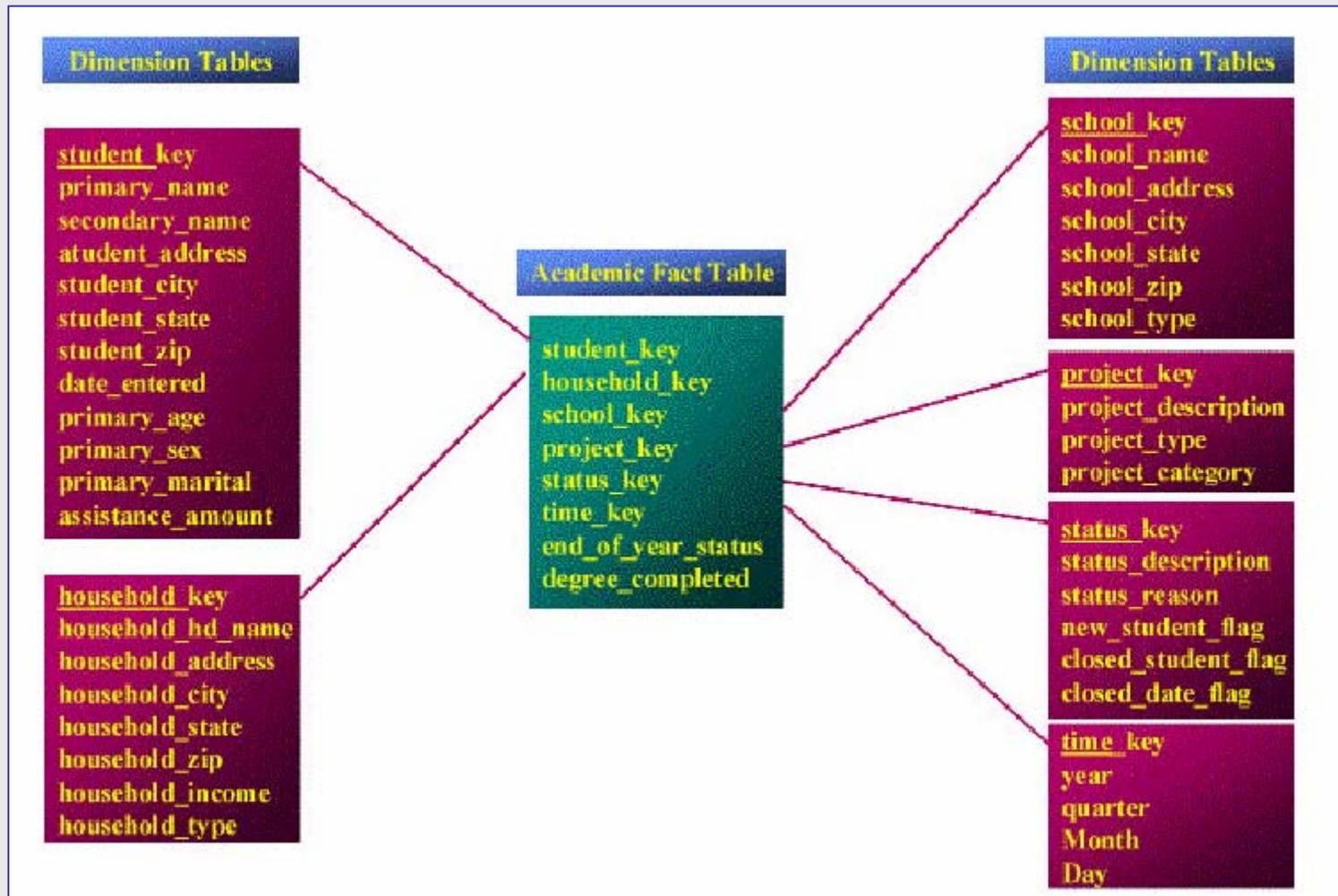
DATAWAREHOUSE: Colección de objetos

- **Orientada al sujeto:**
 - Organizada en torno a los datos más importantes de la empresa.
 - Es bueno para realizar filtros y eliminar información poco importante.
- **Unificada:**
 - Basada en unión de información de varias fuentes.
 - Asegura la consistencia de la información.
- **Variante en el tiempo**
 - Guarda información a través del tiempo.
 - Posee actualizaciones temporales agregadas: no hay actualizaciones diarias.

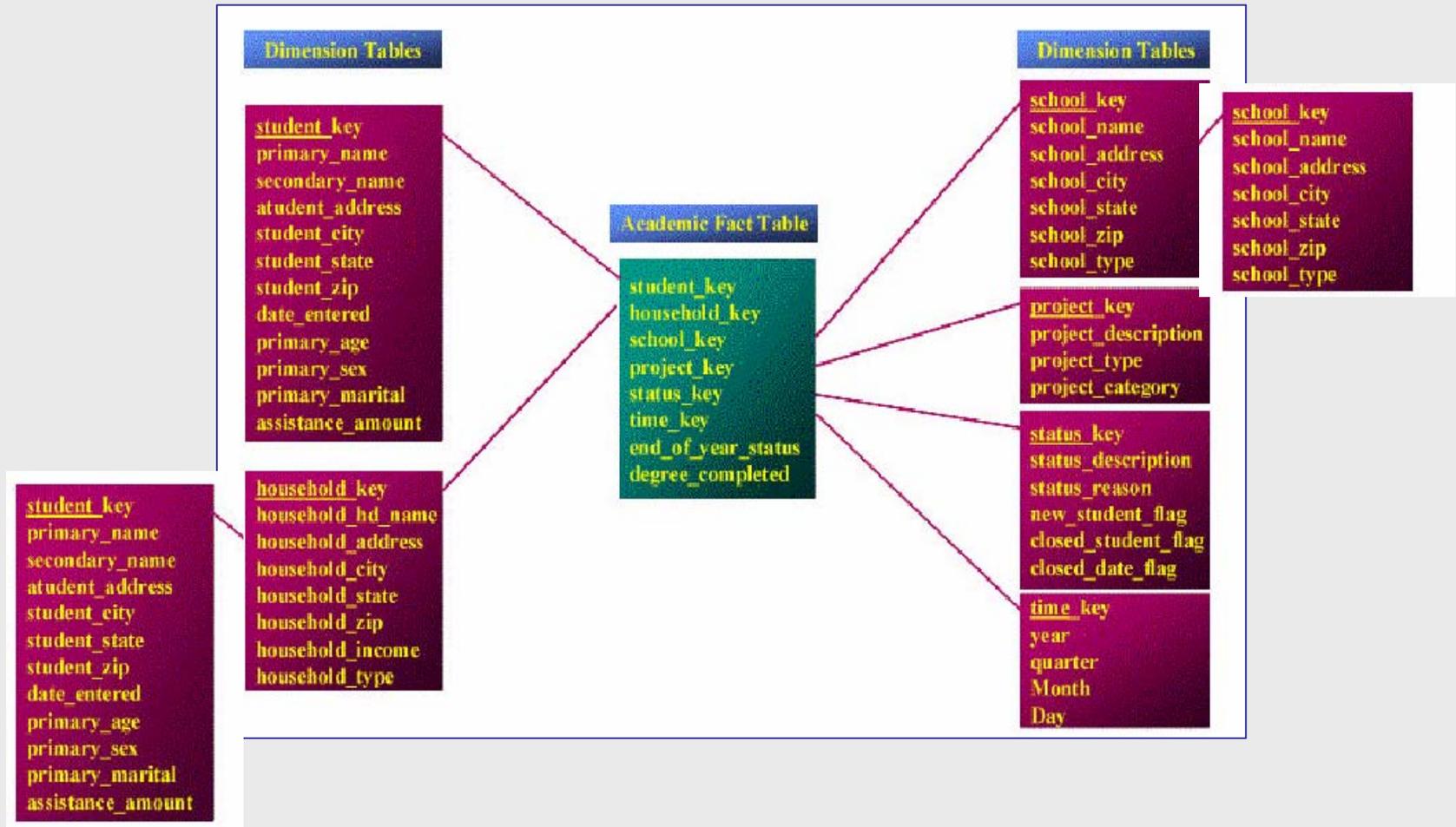
GRÁFICAMENTE ...



ESQUEMA “ESTRELLA”

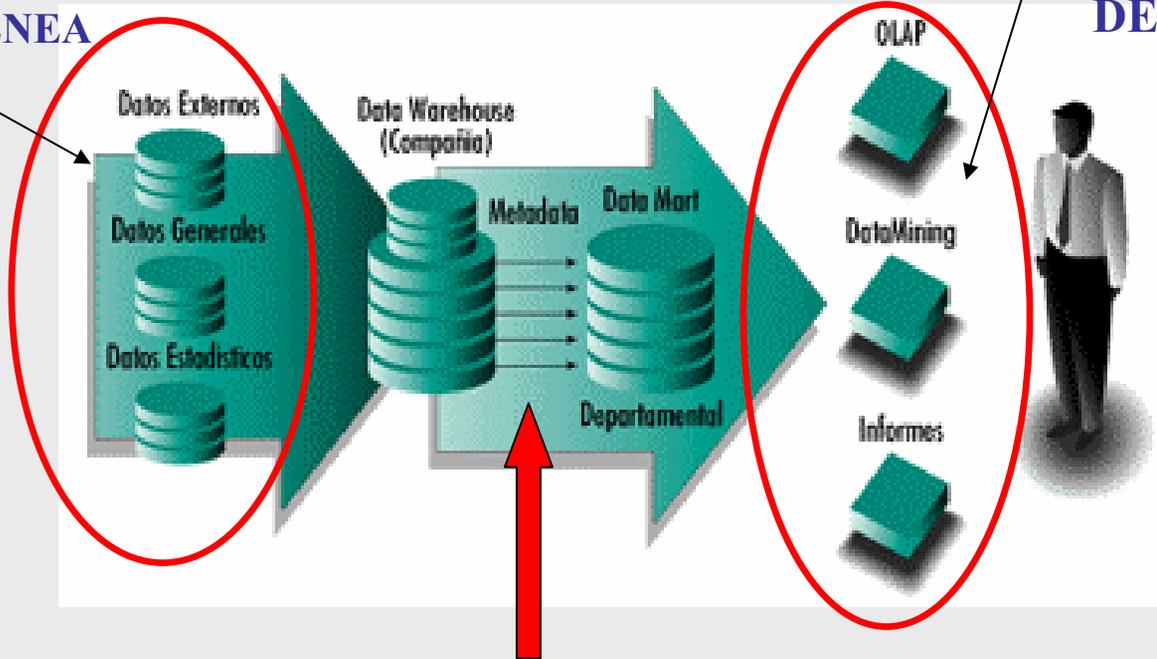


ESQUEMA “COPO DE NIEVE”



ARQUITECTURA MULTICAPAS

INFORMACION
HETEROGENEA



HERRAMIENTAS
DE ANALISIS

METADATOS



Cubos Multidimensionales

NOCIONES BÁSICAS

JAIME MIRANDA

Departamento de Ingeniería Industrial
Universidad de Chile

ALGUNAS NOCIONES...

Consiste en una representación multidimensional de datos de detalle y resumen.

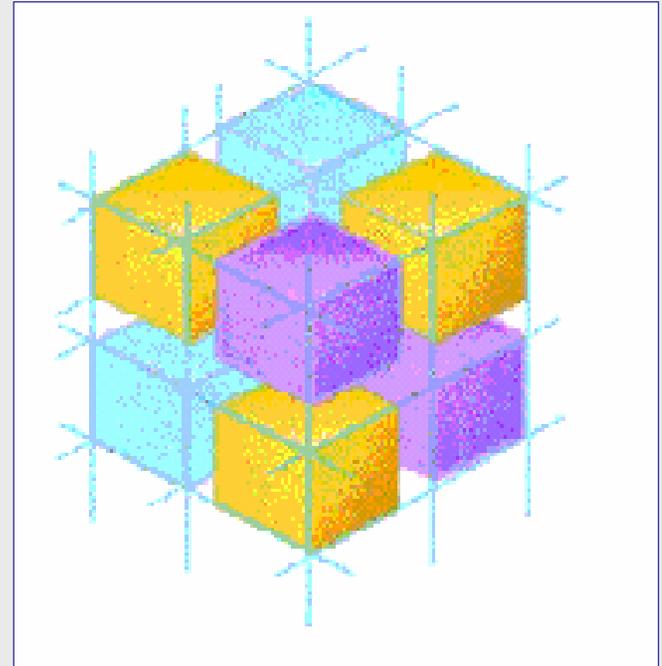
Tiene como objetivo mejorar el rendimiento empresarial en línea y mejorar el rendimiento de las consultas.



ALGUNAS NOCIONES...

Son un subconjunto de datos de la base de datos original.

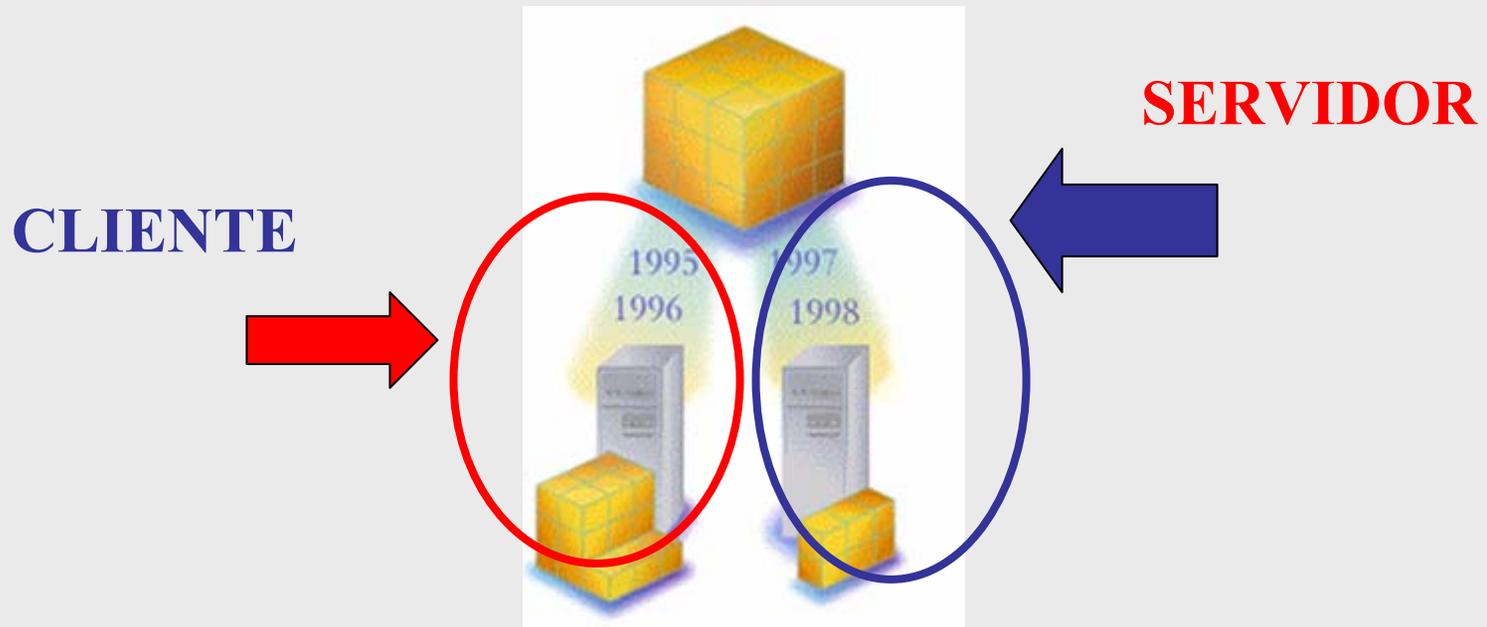
Son capaces de administrar de forma rápida y eficiente grandes cantidades de información.



COMPONENTES DE UN CUBO

ORIGEN DE LOS DATOS

- Identifica y conecta donde se encuentra el almacén de datos la información relevante para resolver un problema.



COMPONENTES DE UN CUBO (2)

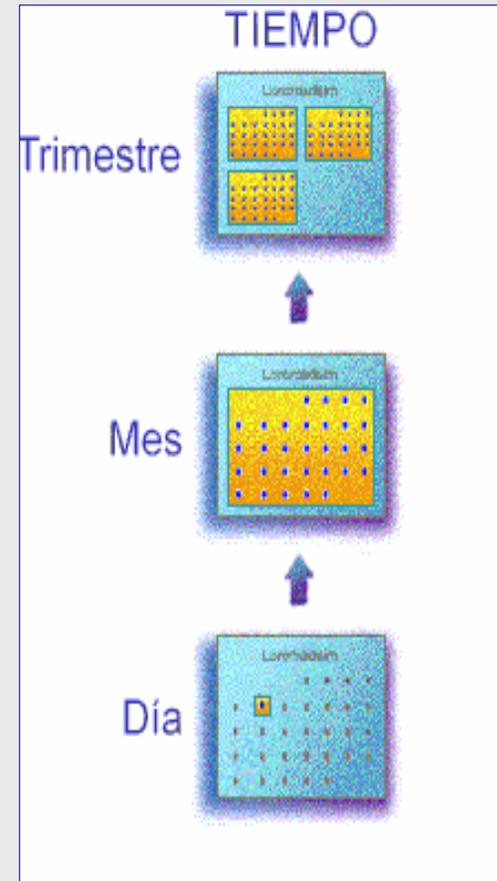
MEDIDAS

- Datos numéricos de interés para los usuarios.
- Que queremos medir o seleccionar.
- Algunos ejemplos:
 - Ventas.
 - Costos.
 - Unidades vendidas.
- Se pueden crear algunas medias:
 - $\text{Beneficios} = \text{Ventas} - \text{Costos}$

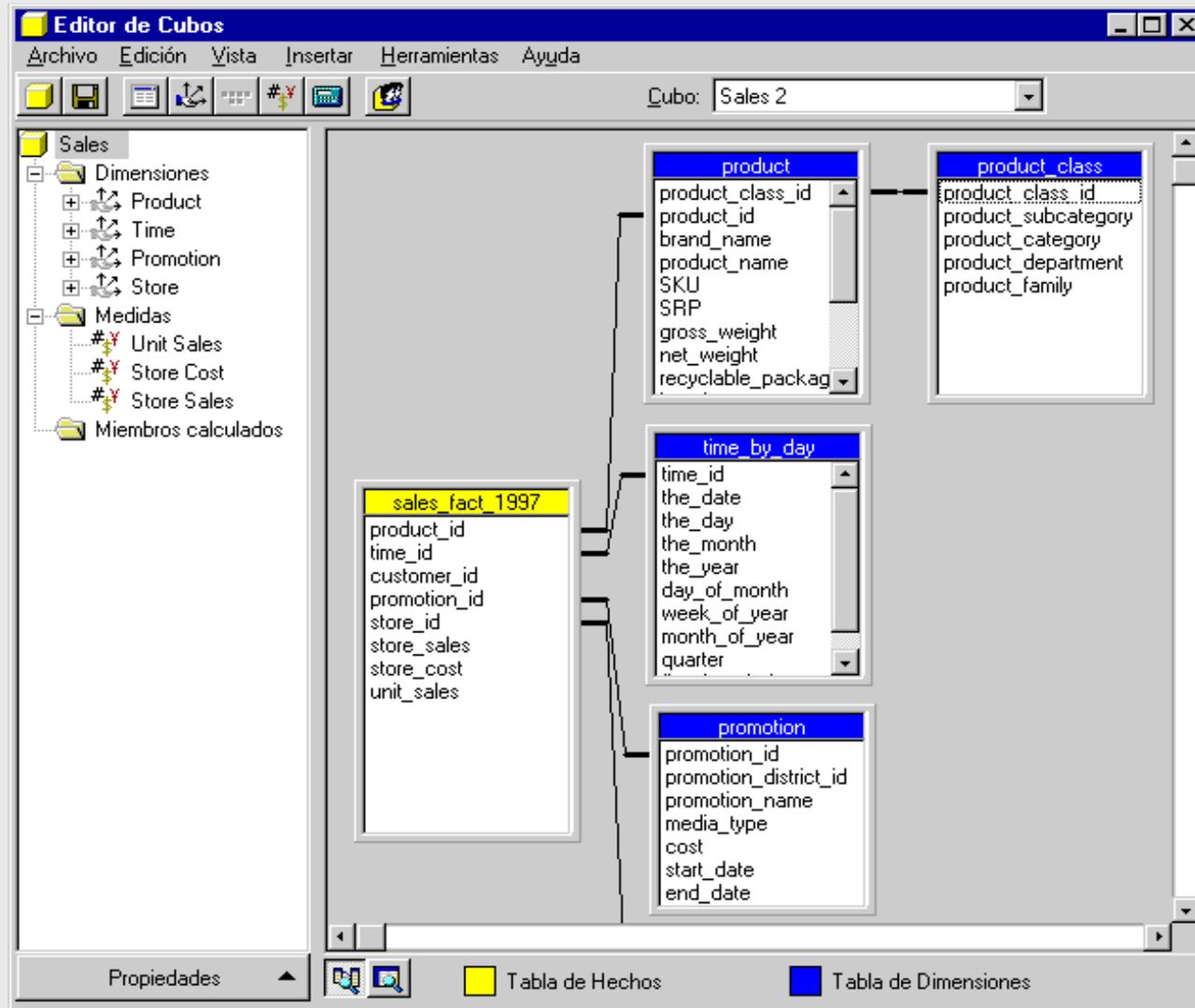


DIMENSIONES

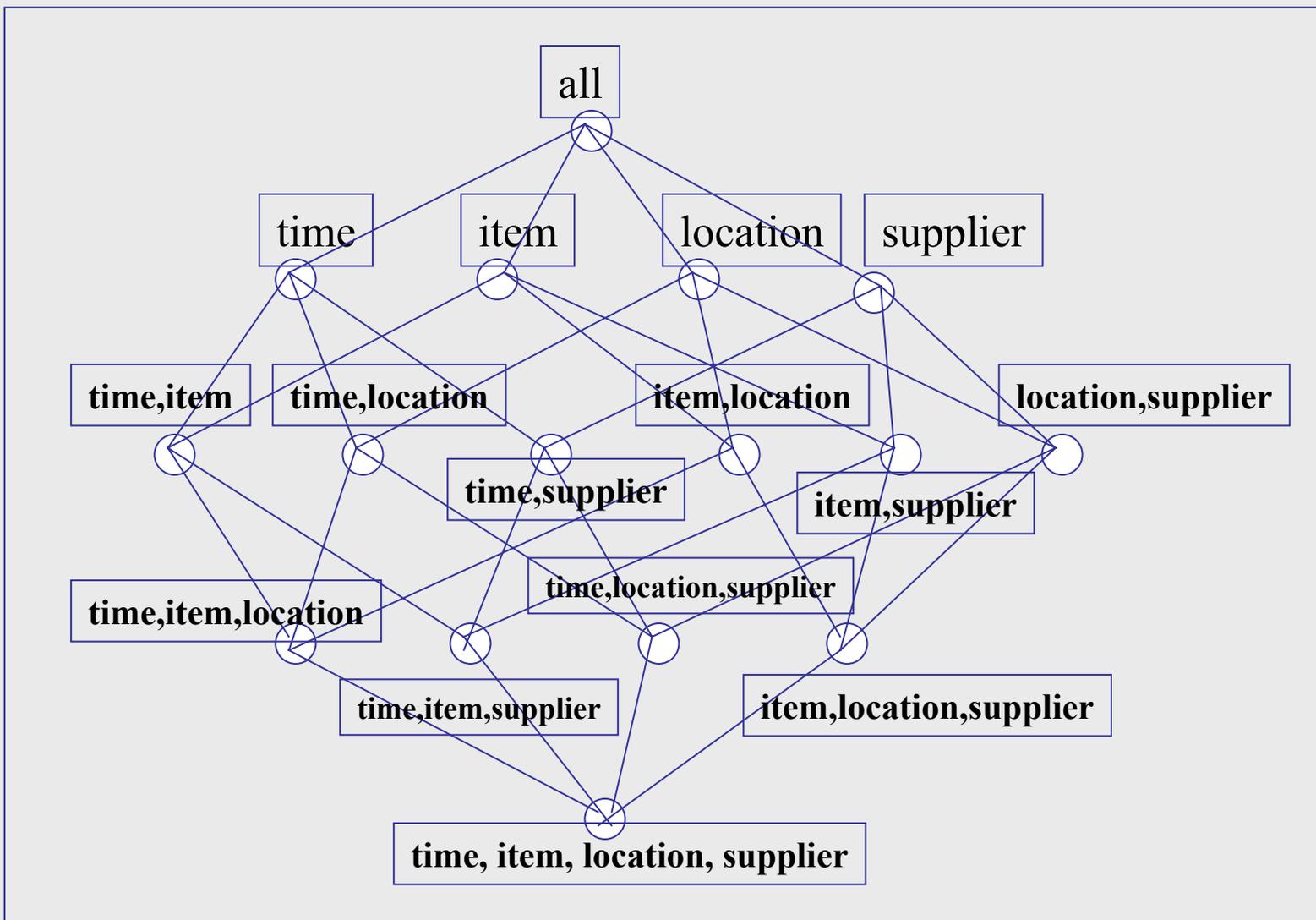
- Representan columnas que describen las categorías a través de las cuales se separan las medidas.
- Similitud con los ejes de un sistema cartesiano.
- Tienen un límite máximo de 64 dimensiones.



EJEMPLO DE UN CUBO



OTRO EJEMPLO DE UN CUBO ...

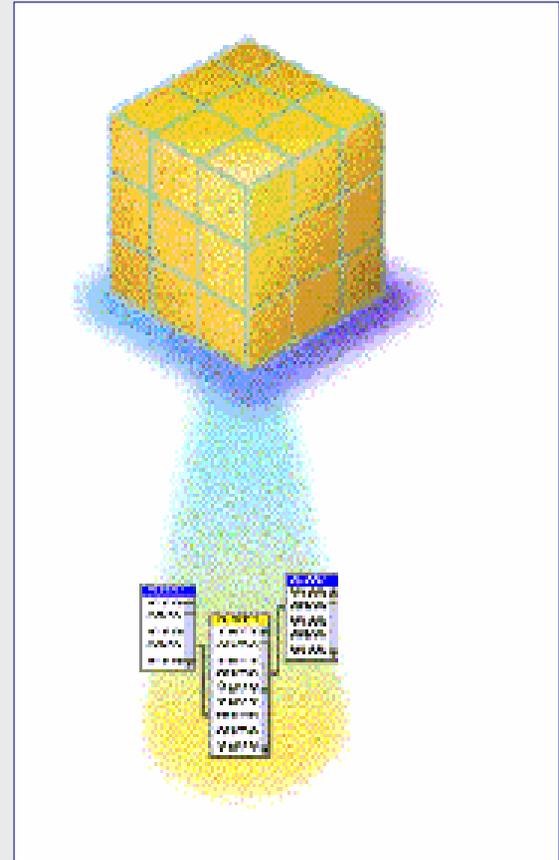


MOLAP (OLAP multidimensional)

- Formato de almacenamiento de alto rendimiento.
- Esta altamente especializado a datos multidimensionales.
- Se aconseja para conjuntos de datos pequeños o medios.
- Es recomendable para cubos de uso frecuente , pues presenta tiempos de respuesta rápidos y eficientes.

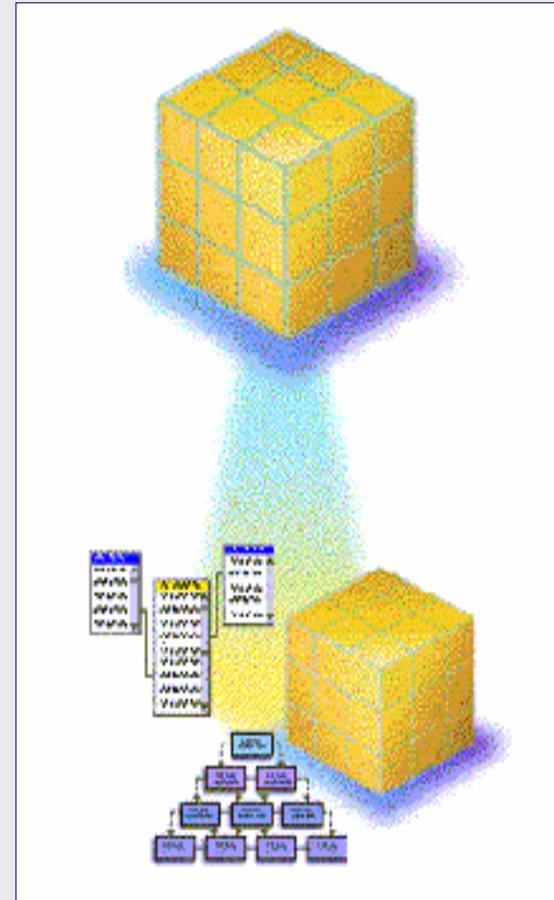
ROLAP

- Los datos permanecen en las tablas originales.
- Se utiliza un conjunto separado de tablas relacionales para hacer referencia a los datos agregados.
- Ideal para bases grandes o datos antiguos que se consultan con poca frecuencia.



HOLAP

- Combinación de ambos modos (OLAP y ROLAP).
- Mantiene los datos originales en tablas relacionales (ROLAP).
- Mantiene los datos agregados en formato multidimensional (MOLAP)



Esto implica:

- Leer las tablas de dimensiones para llenar los miembros con los datos actuales.
- Leer la tabla de hechos.
- Almacenar los datos en el cubo.
- Se debe procesar un cubo cada vez que se ingresen nuevos valores o cuando se modifiquen alguna dimensión o medida.

Ventajas:

- Mejora la eficiencia de las consultas
- Reducen los tiempos de respuesta.



Cubos Multidimensionales

NOCIONES BÁSICAS

JAIME MIRANDA

Departamento de Ingeniería Industrial
Universidad de Chile