

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.

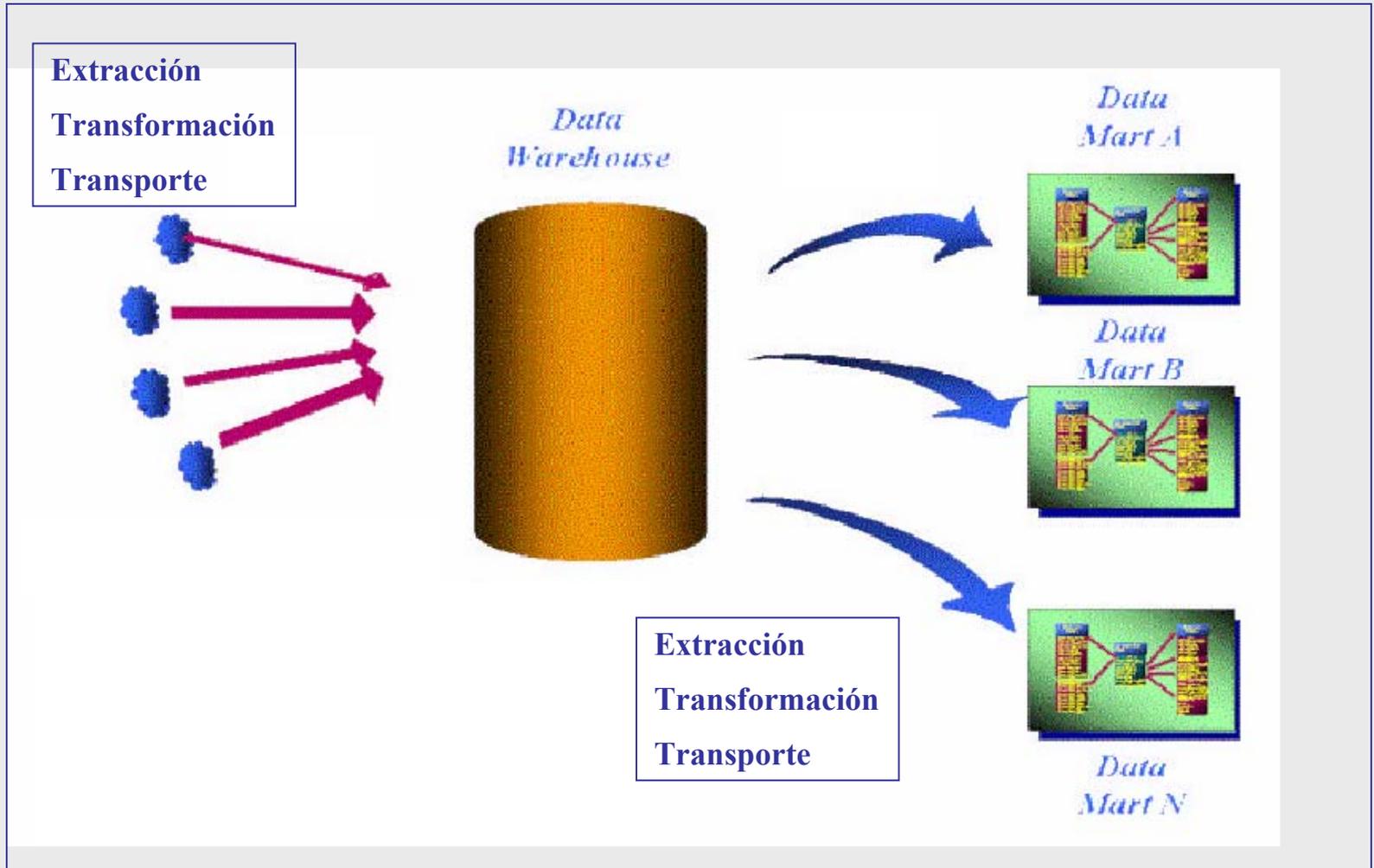
Maneja elevados volúmenes de información.

Permite el mejor funcionamiento de los métodos de Data Mining.

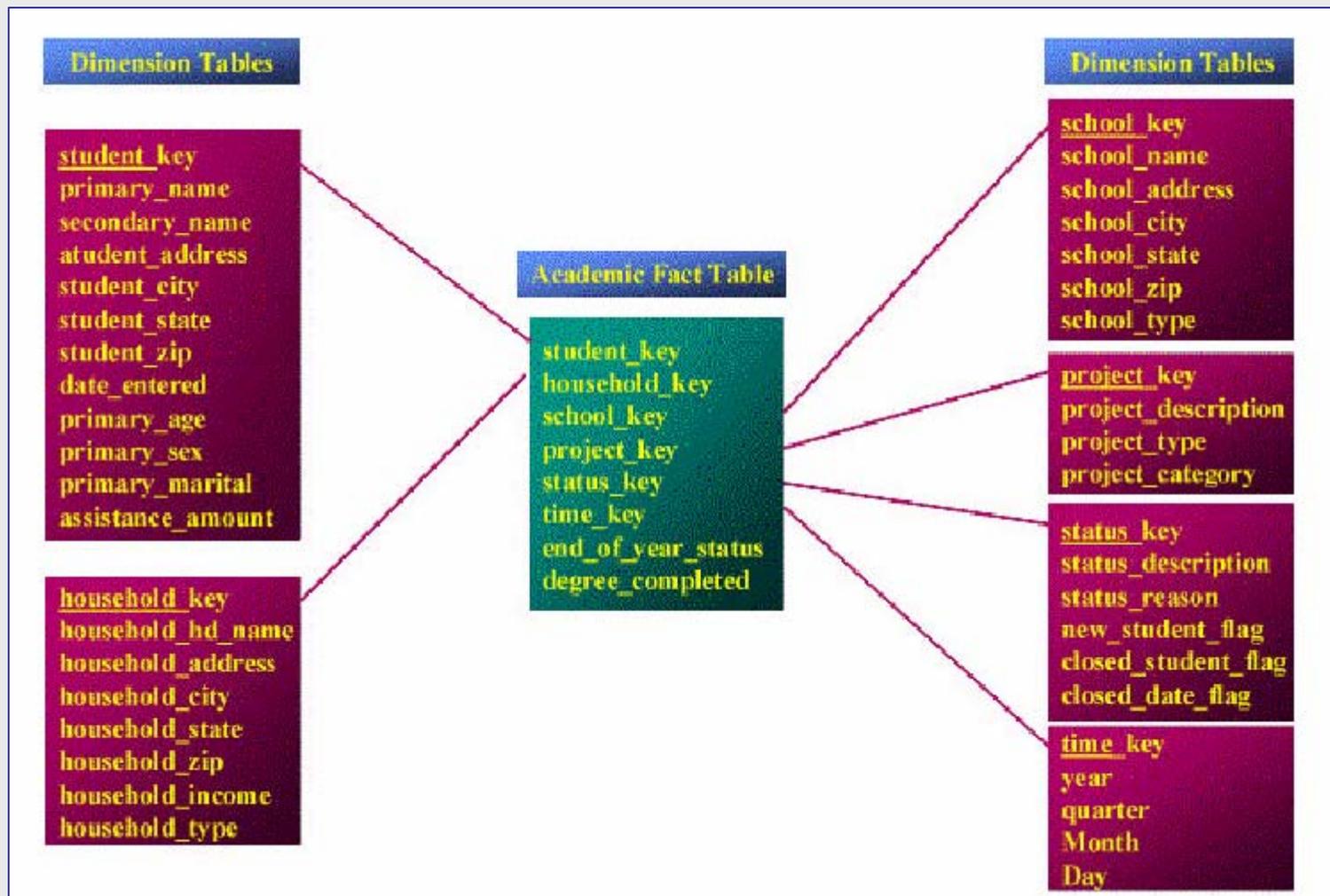
DATAWAREHOUSE: Colección de objetos

- **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.

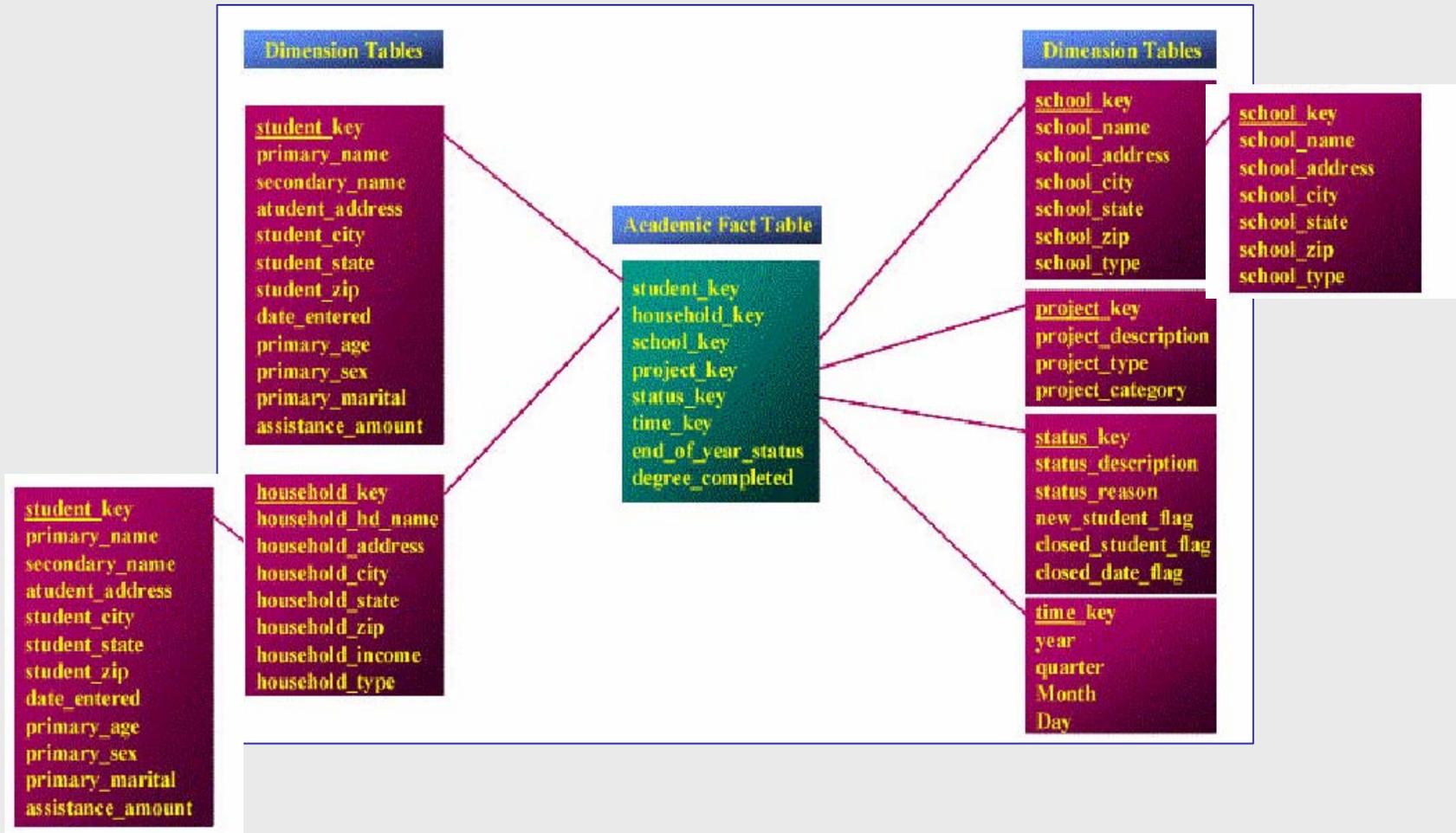
GRÁFICAMENTE ...



ESQUEMA “ESTRELLA”

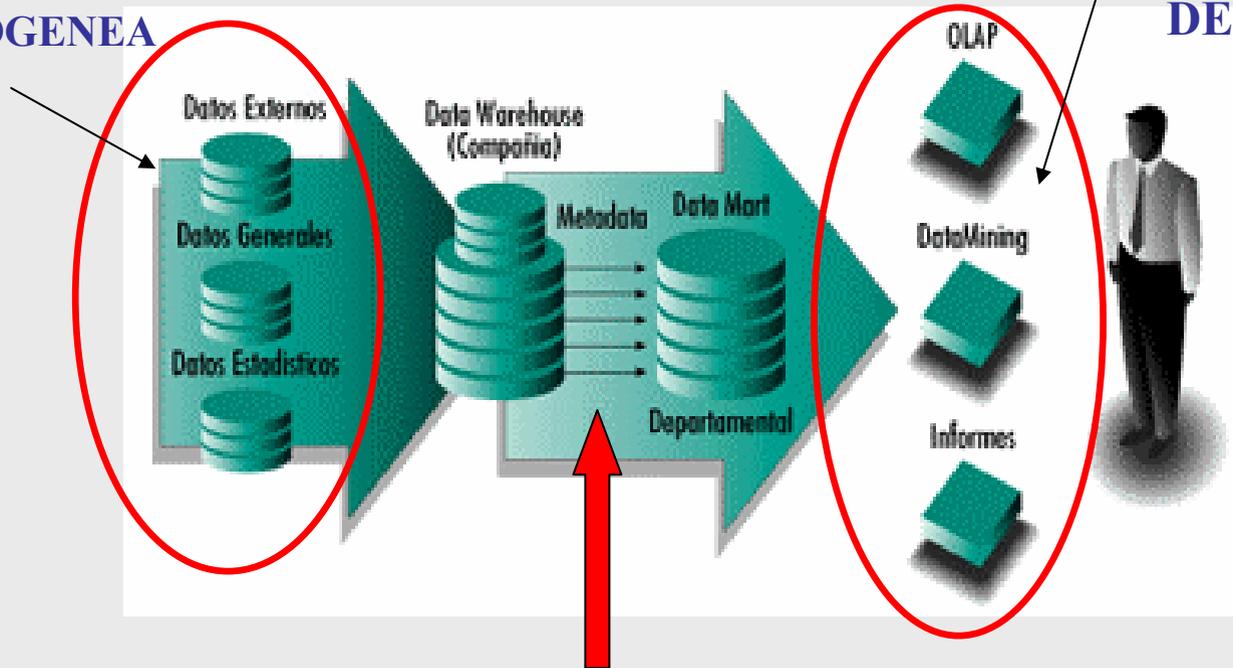


ESQUEMA “COPO DE NIEVE”



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.

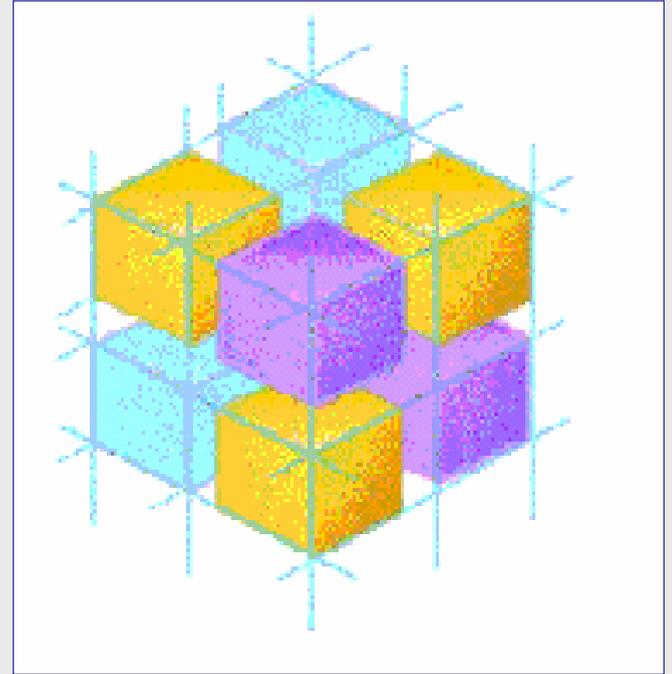
Tiene como objetivo mejorar el rendimiento empresarial en línea y mejorar el rendimiento de las consultas.



ALGUNAS NOCIONES...

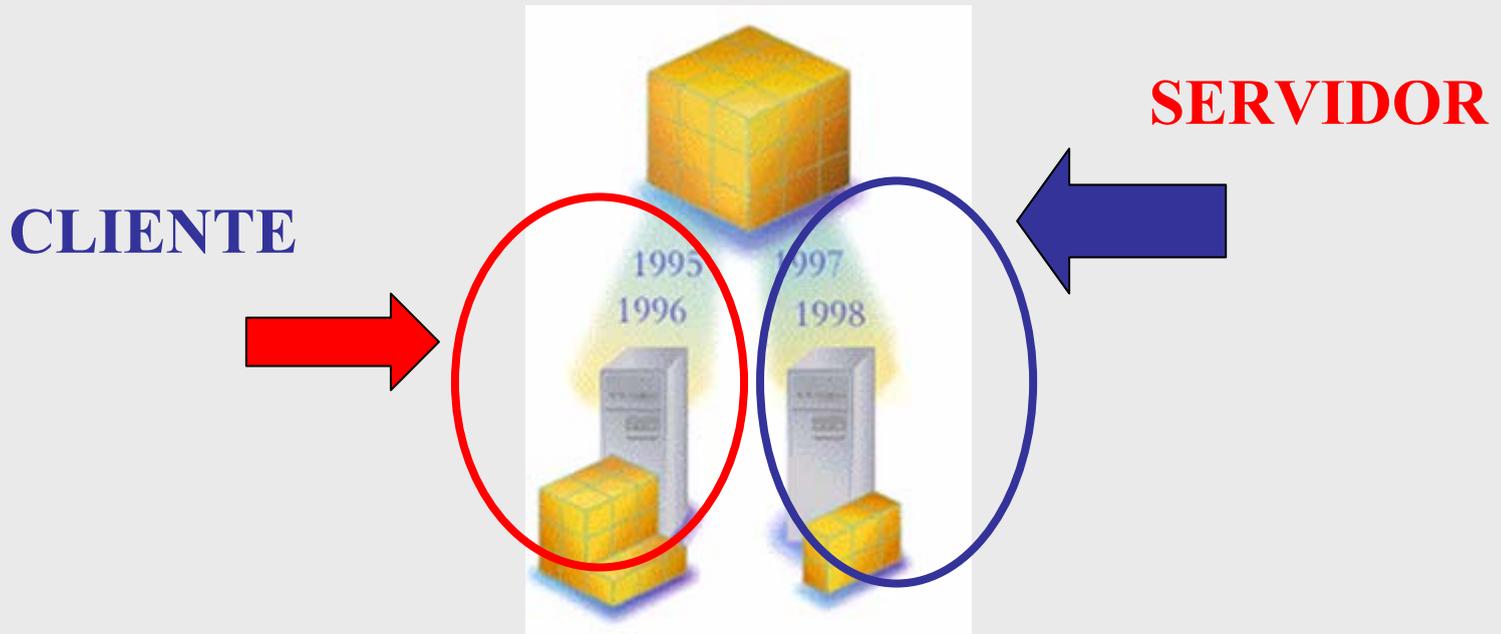
Son un subconjunto de datos de la base de datos original.

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



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

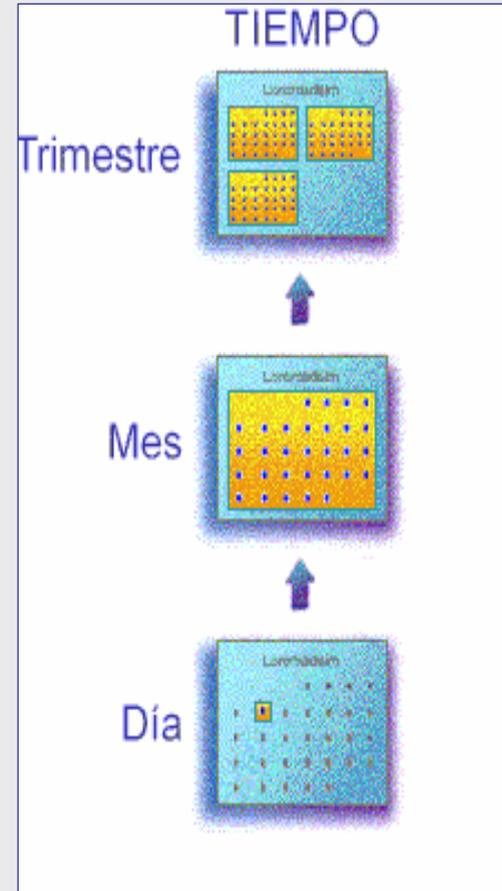
- Datos numéricos de interés para los usuarios.
- Que queremos medir o seleccionar.
- Algunos ejemplos:
 - Ventas.
 - Costos.
 - Unidades vendidas.
- Se pueden crear algunas medias:
 - $\text{Beneficios} = \text{Ventas} - \text{Costos}$



Componentes de un cubo (3)

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

Editor de Cubos

Archivo Edición Vista Insertar Herramientas Ayuda

Cubo: Sales 2

Sales

- Dimensiones
 - Product
 - Time
 - Promotion
 - Store
- Medidas
 - Unit Sales
 - Store Cost
 - Store Sales
- Miembros calculados

sales_fact_1997 (Tabla de Hechos)

- product_id
- time_id
- customer_id
- promotion_id
- store_id
- store_sales
- store_cost
- unit_sales

product (Tabla de Dimensiones)

- product_class_id
- product_id
- brand_name
- product_name
- SKU
- SRP
- gross_weight
- net_weight
- recyclable_packag

product_class (Tabla de Dimensiones)

- product_class_id
- product_subcategory
- product_category
- product_department
- product_family

time_by_day (Tabla de Dimensiones)

- time_id
- the_date
- the_day
- the_month
- the_year
- day_of_month
- week_of_year
- month_of_year
- quarter

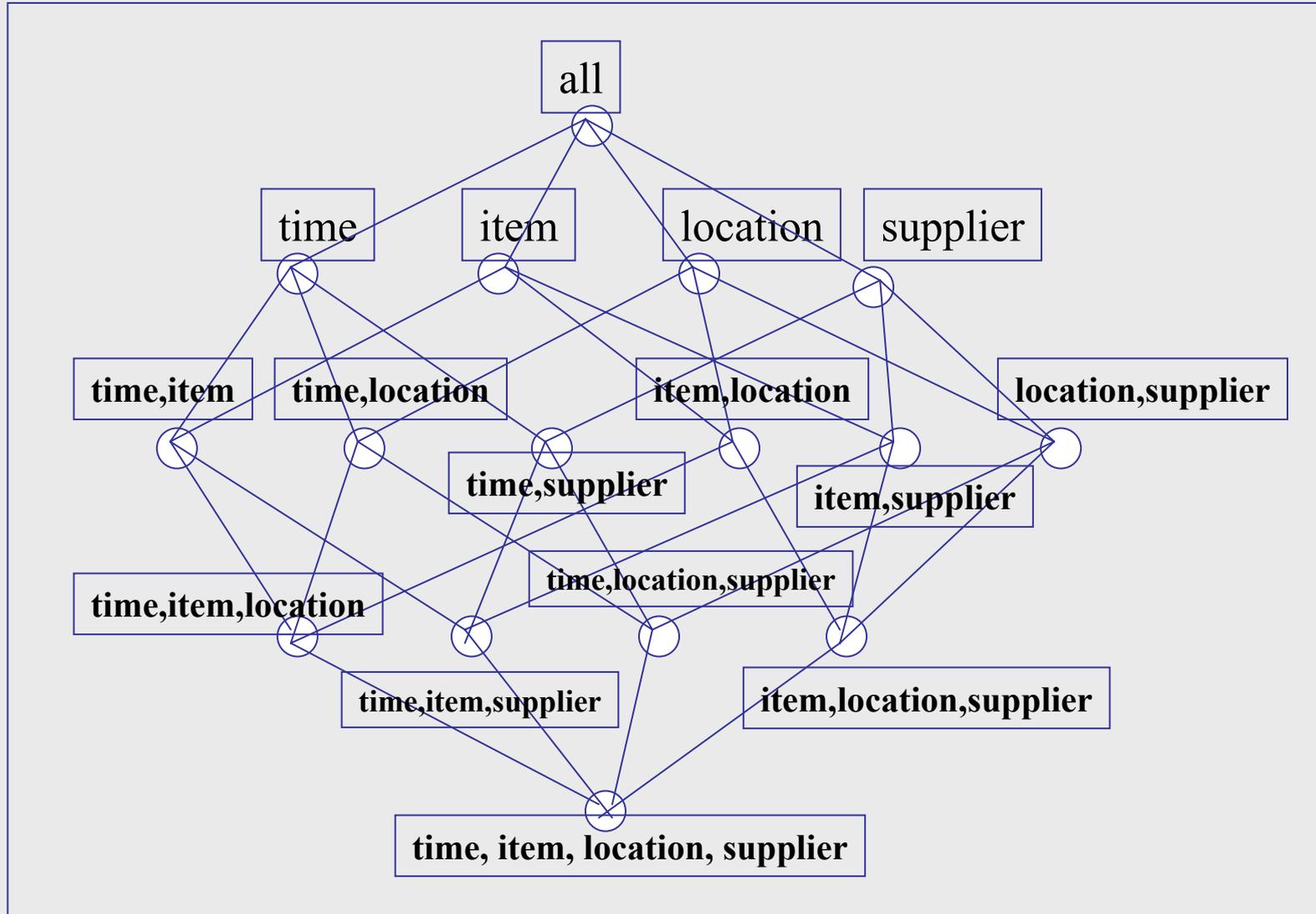
promotion (Tabla de Dimensiones)

- promotion_id
- promotion_district_id
- promotion_name
- media_type
- cost
- start_date
- end_date

Propiedades

Tabla de Hechos Tabla de Dimensiones

OTRO EJEMPLO DE UN CUBO ...

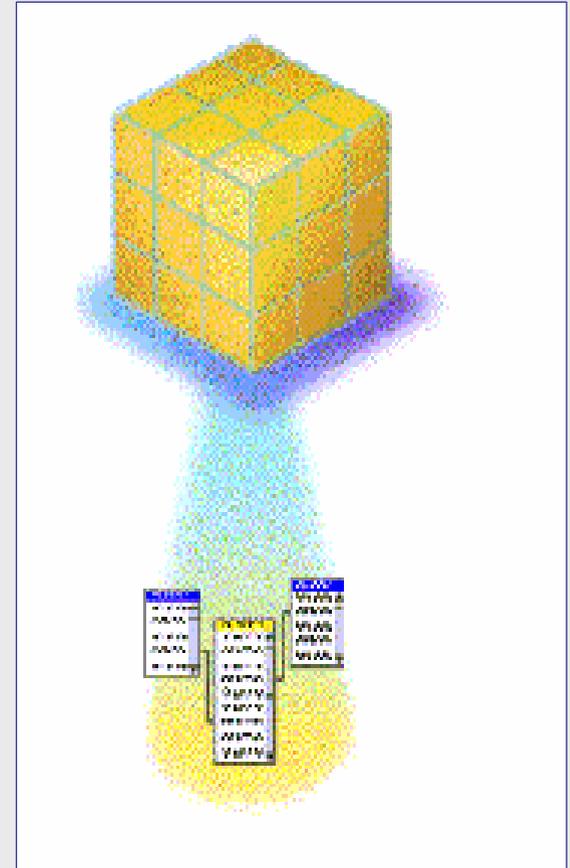


MOLAP (OLAP multidimensional)

- 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.

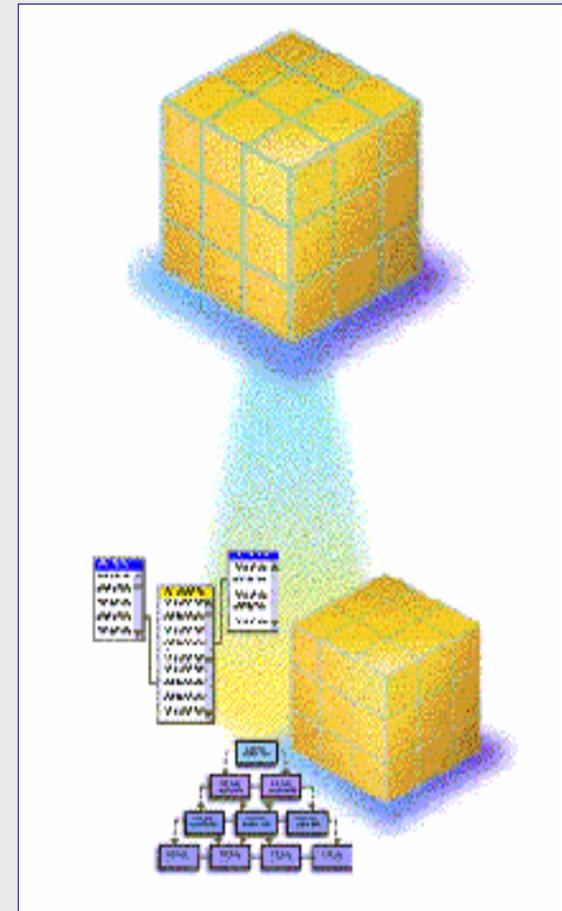
ROLAP

- 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.



Presentación de la Trabajo N°1

CONSTRUCCIÓN DE CUBOS OLAP

JAIME MIRANDA

Departamento de Ingeniería Industrial
Universidad de Chile

INTRODUCCIÓN

Ustedes formaran parte de un equipo de consultor los cuales han sido contratados por el departamento de marketing de la empresa Food Mart.

El departamento de Marketing desea analizar las ventas de su empresa y la forma de operar de las distintas bodegas que esa posee.

Se desea que ustedes realizasen un estudio crítico y den mejoras a las políticas de Marketing.

Ejemplo base del tutorial del SQL Server 7.0

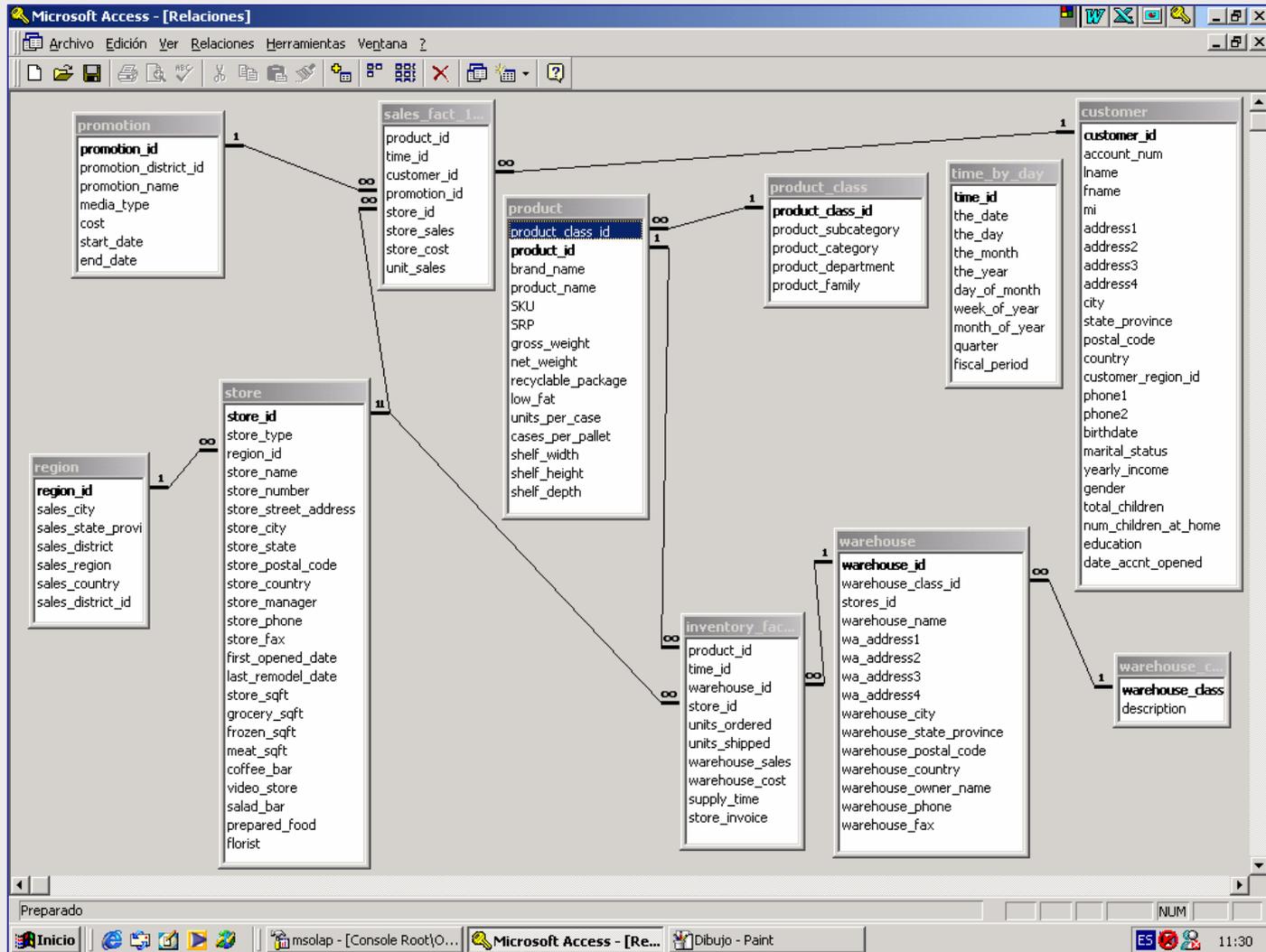
Corresponde a una cadena de supermercado de E.E.U.U.

Enfocada al área de marketing.

Tamaño: 23 Mb.

Base de datos relacional

RELACIONES FOOD-MART DB



Sales_fact_ 1997 y1998:

- Posee entre sus principales registros las transacciones realizadas en cada año, montos de venta, unidades vendidas, costos asociados, etc.

Product:

- Identifica los productos de acuerdo a su nombre, dimensiones físicas y de almacenamiento, entre otras.

Product_Class:

- Clasificación según categoría, departamento y familia

Promotion:

- Identificación de las promociones, duración, medio de comunicación y costo.

Warehouse:

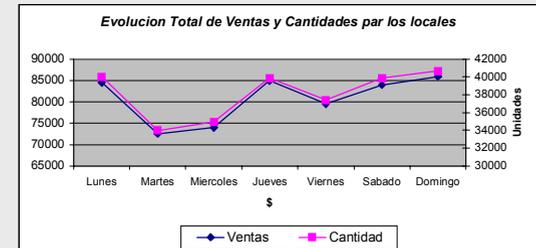
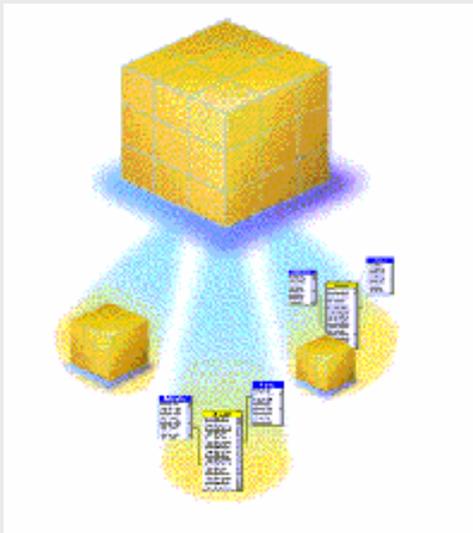
- Identificación de las distintas bodegas, nombre, dirección, país, teléfono, etc.

Time_by_Day:

- Identificadores diarios, donde para cada día señala la semana, mes, trimestre, año.

OBJETIVO GENERAL

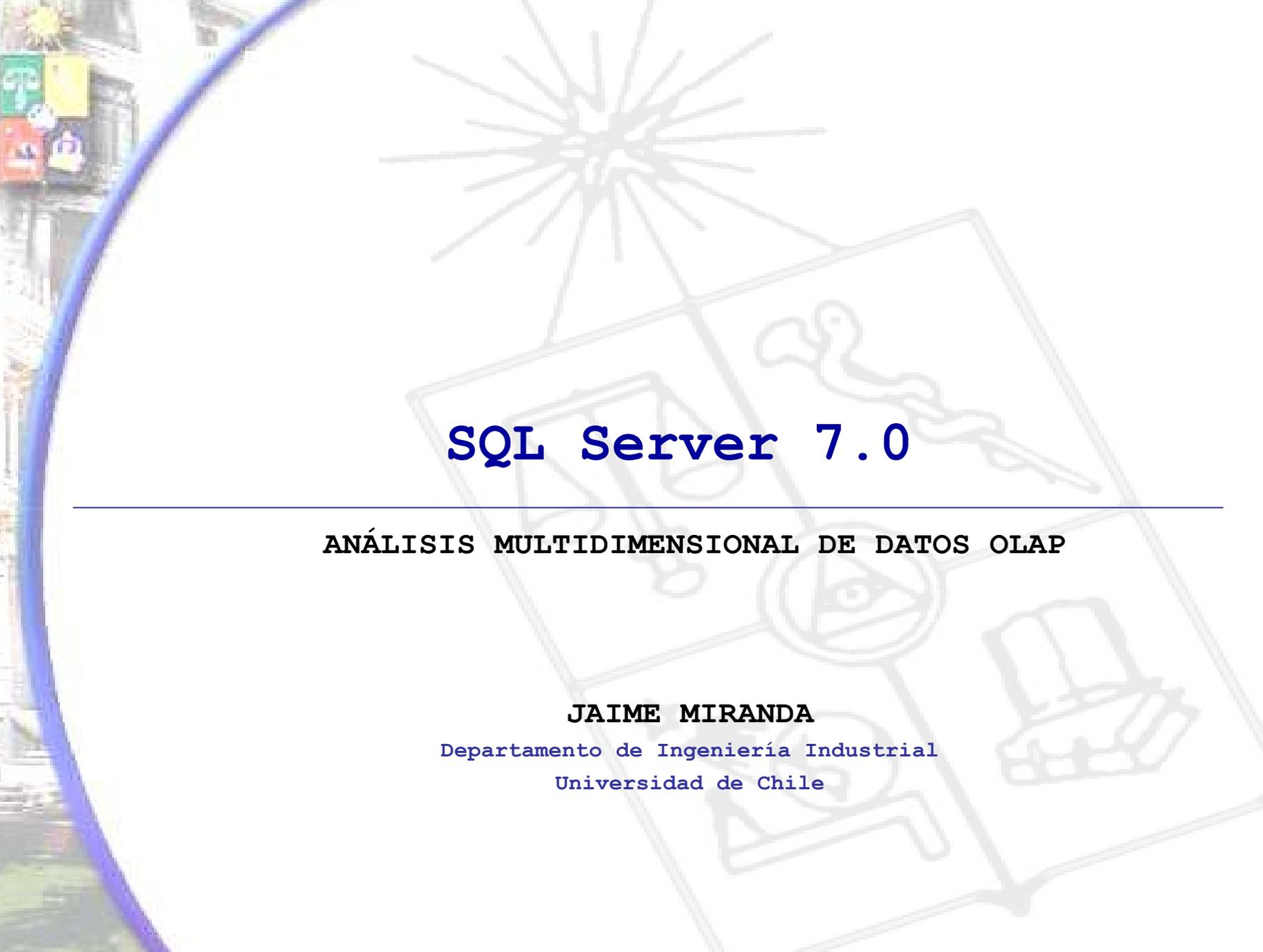
- “Aprender a usar las herramientas OLAP, con el fin de hacer análisis multidimensional de los datos aplicando esto al análisis de patrones de comportamiento de alguna variable específica.”



OBJETIVOS ESPECIFICOS

- Diseñar y construir cubos multidimensionales con SQL Server.
- Manejar y aprovechar la flexibilidad de los cubos para el análisis.
- Analizar la información para responder consultas a través de un software estadístico.
- Buscar patrones de comportamiento y dar soluciones a los problemas de la empresa con una actitud activa hacia las políticas de marketing.

Determine los beneficios (ventas-costos) y unidades vendidas durante el primer trimestre del año 1997 en las tiendas de México. ¿Esto es igual en todos los países?...



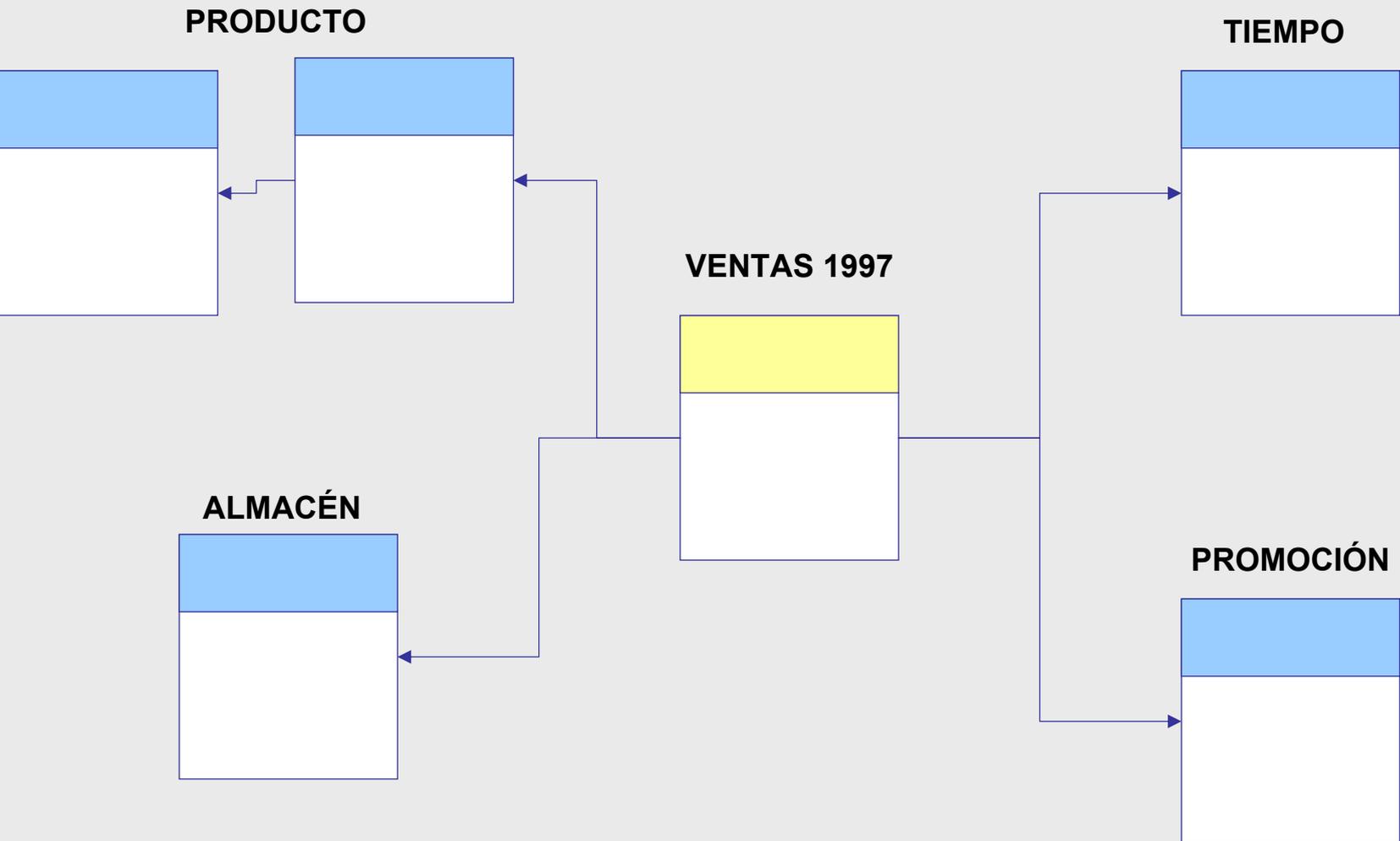
SQL Server 7.0

ANÁLISIS MULTIDIMENSIONAL DE DATOS OLAP

JAIME MIRANDA

Departamento de Ingeniería Industrial
Universidad de Chile

CUBO A CONSTRUIR ...



Configurar la conexión con el origen de datos del sistema

1. En el menú Inicio de Windows, seleccione Configuración, haga clic en Panel de Control y, después, haga doble clic en Herramientas Administrativas y luego en Orígenes de datos ODBC.
2. En la ficha DSN de sistema, haga clic en Agregar.
3. Seleccione Controlador de Microsoft Access (*.mdb) y haga clic en Finalizar.
4. Escriba "Tutorial_IN78J" en el cuadro Data Source Name (Nombre de origen de datos) y, a continuación, haga clic en Select (Seleccionar), bajo Database (Base de datos).

Configurar la conexión con el origen de datos del sistema

5. En el cuadro de diálogo Select Database (Seleccionar base de datos), explore la carpeta C:\Archivos de programa\OLAP Services\Samples y seleccione FoodMart.mdb. Haga clic en Aceptar.
6. Haga clic en Aceptar en el cuadro de diálogo ODBC Microsoft Access Setup (Configuración de ODBC para Microsoft Access).
7. Haga clic en Aceptar en el cuadro de diálogo Administrador de orígenes de datos ODBC.

Configurar la base de datos y el origen de datos

1. En la vista de árbol del Administrador de OLAP, expanda Servidores de OLAP.
2. Haga clic en el nombre del servidor. Se establecerá una conexión con el Servidor de OLAP (Aparece en Panel de control /Sistemas).
3. A continuación, haga clic con el botón secundario del mouse (ratón) en el nombre del servidor y seleccione Nueva base de datos en el menú contextual.
4. En el cuadro de diálogo Base de datos escriba el nombre "Tutorial_IN78J" en el cuadro Nombre de la base de datos y, después, haga clic en Aceptar.
5. En la vista de árbol del Administrador de OLAP, expanda el servidor y, a continuación, expanda la base de datos "Tutorial_IN78J" que acaba de crear.

Configurar el origen de datos

1. En la vista de árbol del Administrador de OLAP, expanda la carpeta Biblioteca, haga clic con el botón secundario del mouse en la carpeta Orígenes de datos y, finalmente, seleccione Nuevo origen de datos en el menú contextual.
2. En el cuadro de diálogo Propiedades de Vínculo de datos, ficha Proveedor, seleccione Microsoft OLE DB Provider for ODBC Drivers en la lista.
3. Haga clic en la ficha Conexión, a continuación, en la lista desplegable “Utilizar nombre de origen de datos”, seleccione Tutorial_IN78J” .

Configurar el origen de datos

4. Haga clic en Probar conexión para asegurarse de que todo funciona correctamente. Aparecerá un mensaje en el cuadro de diálogo Vínculo de datos Microsoft, que indica que la conexión se ha establecido con éxito. Haga clic en Aceptar.
5. Haga clic en Aceptar para cerrar el cuadro de diálogo Propiedades de Vínculo de datos.



SQL Server 7.0

ANÁLISIS MULTIDIMENSIONAL DE DATOS OLAP

JAIME MIRANDA

Departamento de Ingeniería Industrial
Universidad de Chile