



Data Warehouses

*Almacenamiento y Recuperación de la
Información*

Manuel Serrano

E.S. Informática CR – UCLM

Manuel.Serrano@uclm.es

<http://alarcos.inf-cr.uclm.es/per/mserrano/>



Introducción

- Datos <-> Información
- OLAP <-> OLTP
- Diagramas en estrella
- Modelo Multidimensional
 - Cubos



Datos vs Información

- Una compañía puede capturar grandes cantidades de datos en su trabajo diario
- Estos datos representan el estado actual del negocio
- Es importante derivar información de estos datos
 - Examinando distintos contextos
 - Determinando las relaciones entre los hechos
 - Comprendiendo como se reflejan los objetivos de la empresa en los datos



Datos vs Información

- Navegar por una gran cantidad de información es complicado
- Es conveniente crear un sistema para el soporte a la toma de decisiones (DSS)
- Crear un DSS es una tarea cara, larga y complicada, pero puede ser amortizada si satisface las necesidades de información



Sistemas OLTP

- On-Line Transactional Processing
- Procesan las transacciones de tiempo real de un negocio
- Contienen estructuras de datos optimizadas para la introducción y la edición de los datos
- Proporcionan capacidades muy limitadas para la toma de decisiones



Características de un DW

- Proporciona datos para el análisis de los procesos de negocio
- Integra datos de fuentes heterogéneas
 - Generalmente los distintos sistemas OLTP de la empresa
- Combina datos fuentes validados
 - Se controla la integridad de los datos a la hora de integrarlos

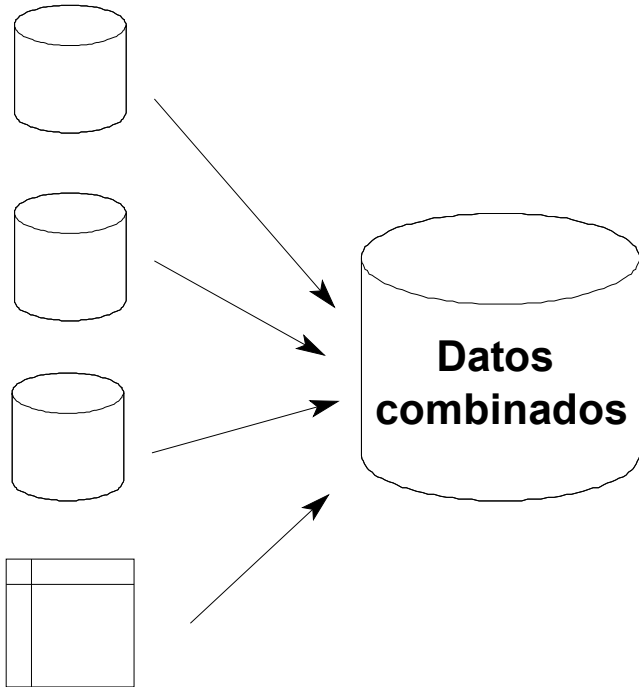


Características de un DW

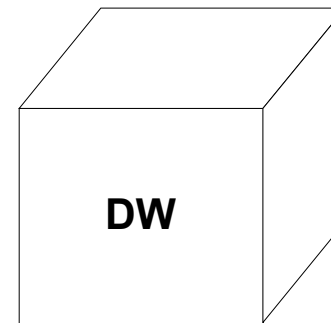
- Organiza los datos en grupos específicos y no volátiles
 - Un DW es estático
 - Las inserciones y las actualizaciones son periódicas
- Almacena los datos en estructuras optimizadas para la distribución de los datos y las consultas
 - Lo más importante son las consultas

Componentes de un DW

Fuentes (OLTP)



Limpieza



Usuarios

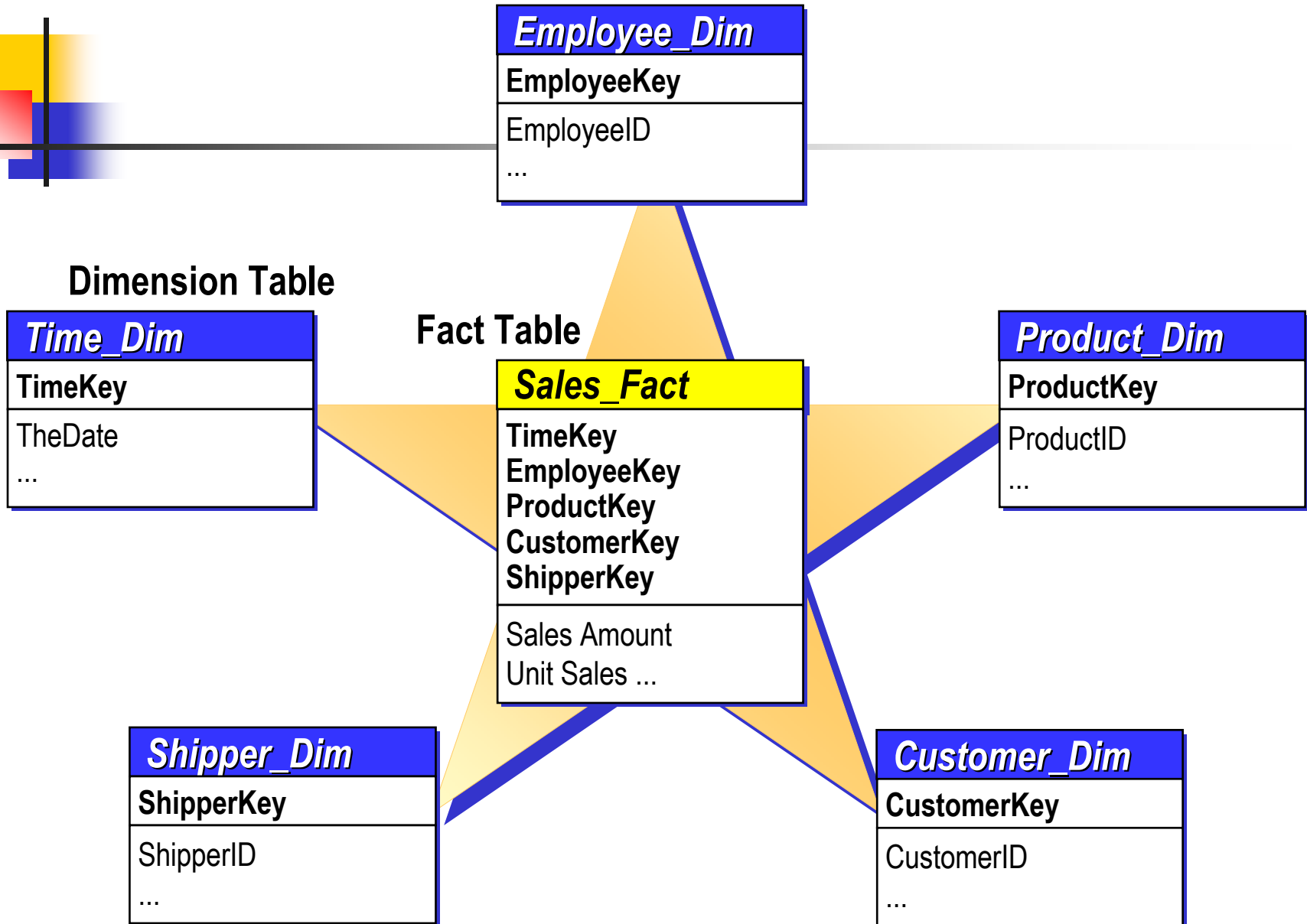




Bases de datos OLAP

- Optimizadas hacia las consultas rápidas
- Incorporan un motor de cálculos para los análisis numéricos
- Se basan en un modelo de datos conceptual e intuitivo
- Permiten la visualización multidimensional de los datos
 - “Drill down” y “Drill up”
 - Pivotar las vistas

Diseño: Esquema en estrella



Diseño: Tabla de hechos

Dimension Tables

customer_dim

201	ALFI	Alfreds

product_dim

25	123	Chai

time_dim


134	1/1/2000


sales_fact Table


<i>customer_key</i>		<i>time_key</i>	<i>quantity_sales</i>	<i>amount_sales</i>
201	25	134	400	10,789

The **grain** of the sales_fact table is defined by the lowest level of detail stored in each dimension

Diseño: Tablas dimensionales

customer *	
	customer_id
	account_num
	lname
	fname
	mi
	address1
	address2
	address3
	address4
	city
	state_province
	postal_code
	country

State *	
	State_ID
	State_Name
	Region
	Country

time_by_day *	
	time_id
	the_date
	the_day
	the_month
	the_year
	day_of_month
	week_of_year
	month_of_year
	quarter
	fiscal_period

- Describen entidades del negocio
- Incorporan el contexto a los datos numéricos
- Presenta los datos organizados en jerarquías



Diseño: Dimensiones OLAP

OLAP

REGION

West

CA

OR

East

MA

NY

Relational

REGION

West

East

STATE REGION

CA West

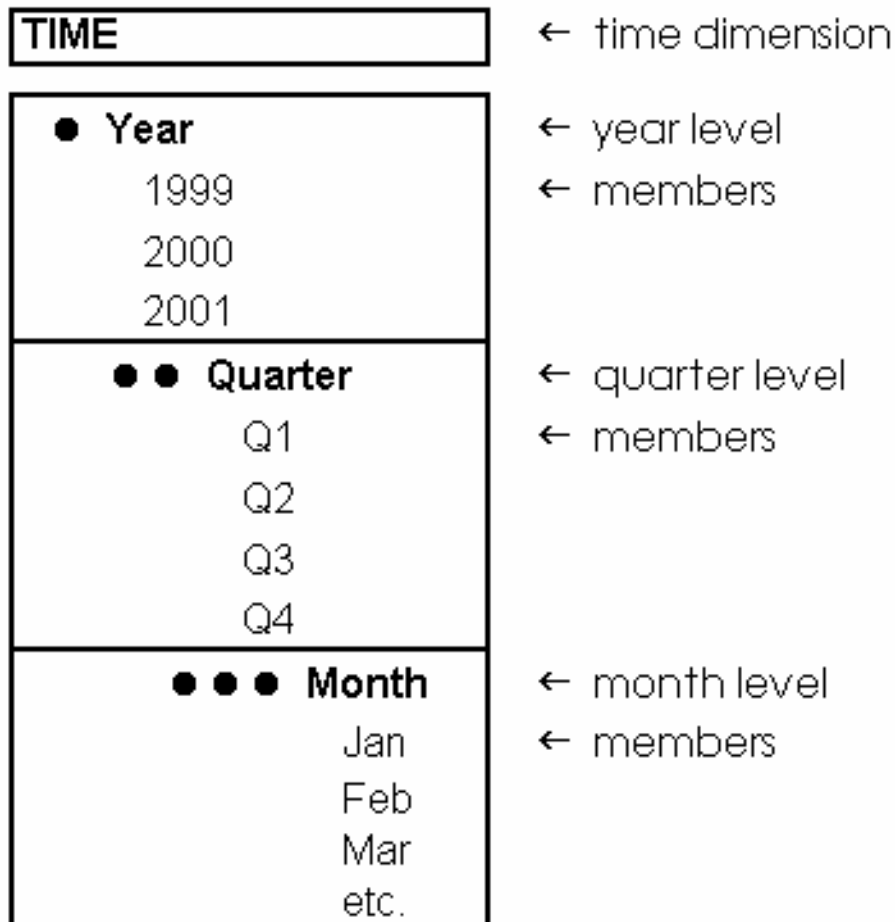
OR West

MA East

NY East



Diseño: Dimensiones OLAP



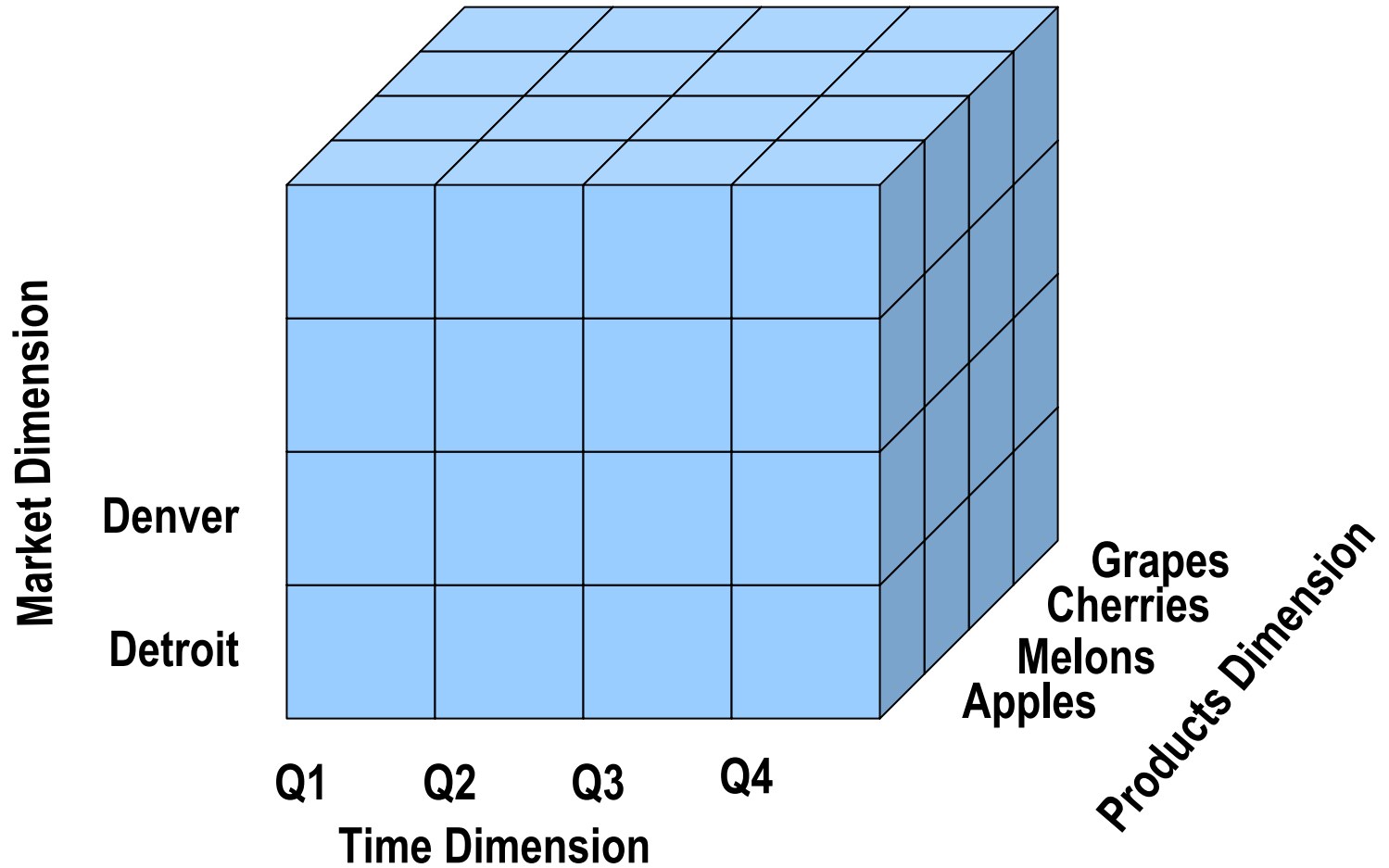


Medidas del Cubo

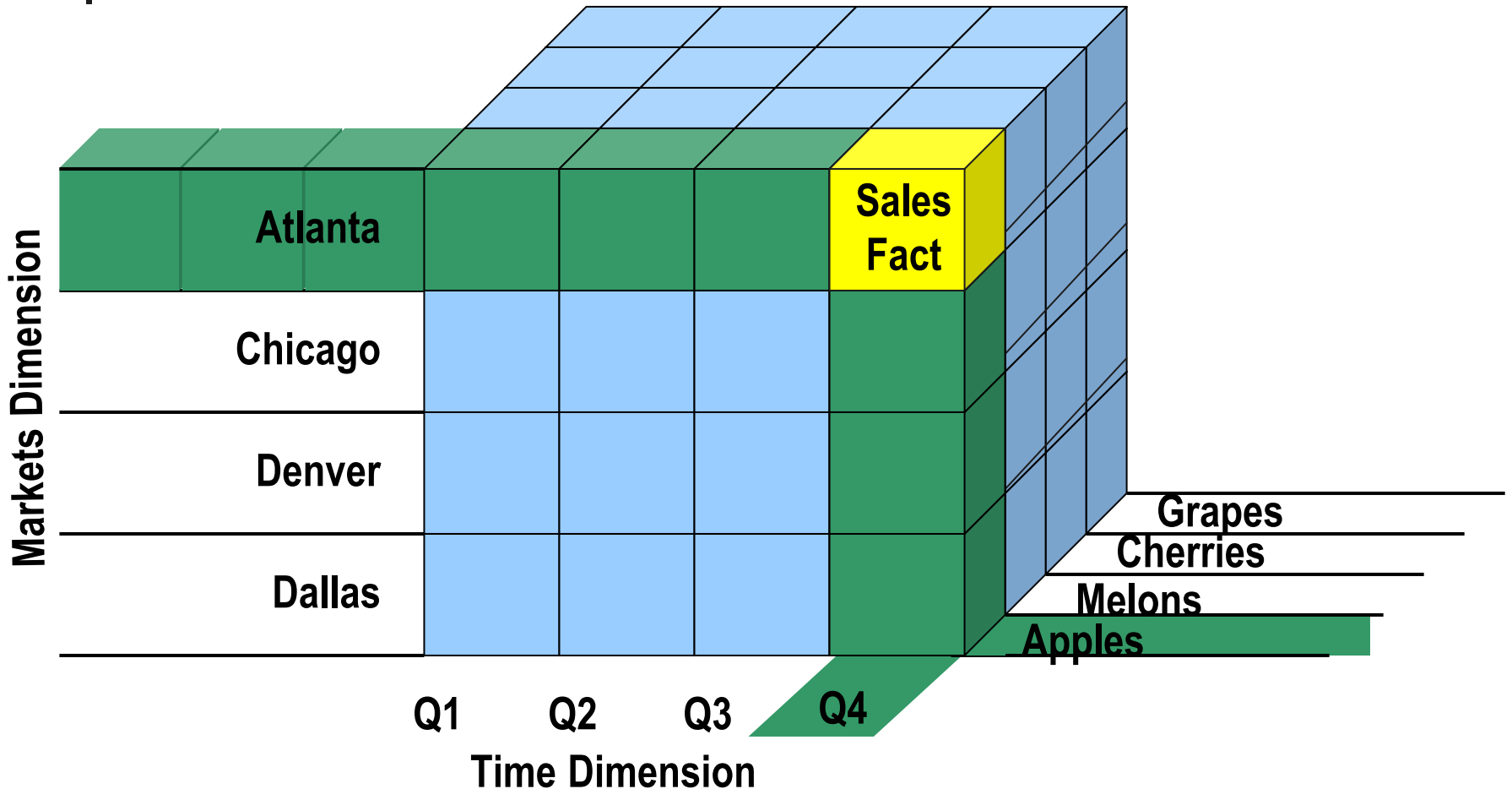
- Son valores de interés
- Son los hechos de las tabla de hechos
- Son la intersección de las dimensiones a todos los niveles
- Se agregan a todos los niveles de detalle
- Forman una dimensión



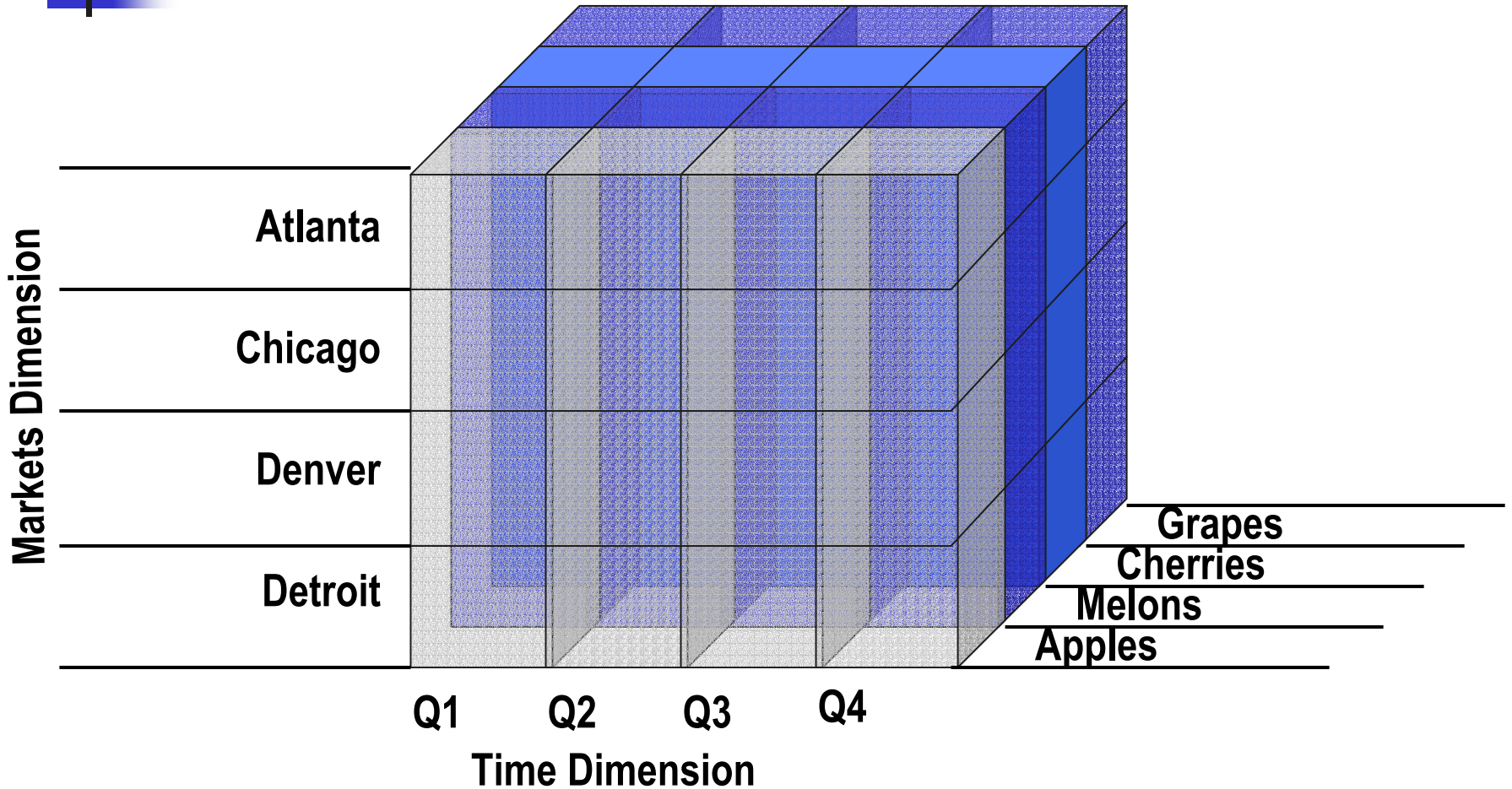
Cubos



Accediendo a los cubos


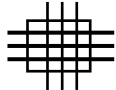




Rodaja de un cubo (Slice)








Dimensiones y jerarquías

- Las dimensiones permiten:
 - Slice 
 - Dice 
- Las jerarquías permiten
 - Drill down 
 - Drill up 

DW Conceptual

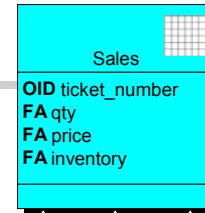
- Perfil UML para modelado conceptual (Trujillo et al., 2001)

Nombre	Descripción	Icono
Fact	Las Clases de este estereotipo representan hechos en un modelo multidimensional	
Dimension	Las Clases de este estereotipo representan dimensiones en un modelo multidimensional	
Base	Las Clases de este estereotipo representan niveles de una jerarquía dimensional en un modelo multidimensional	

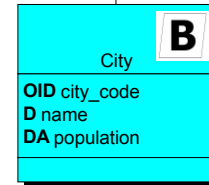
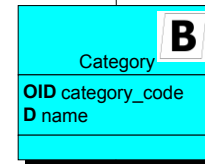
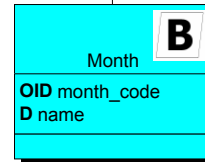
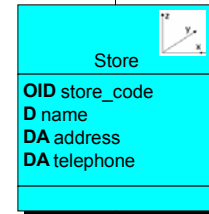
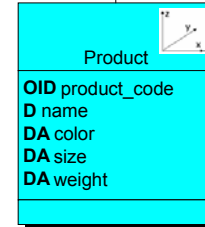
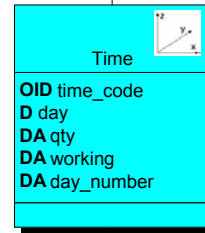
Nombre	Descripción	Icono
OID	Los atributos con este estereotipo representan los OID de clases factuales, dimensionales o base	OID
FactAttribute	Los atributos con este estereotipo representan atributos de clases factuales	FA
Descriptor	Los atributos con este estereotipo representan atributos descriptores de clases dimensionales o base	D
DimensionAttribute	Los atributos con este estereotipo representan atributos de clases dimensionales o base	DA

DW Conceptual

Fact Class



Dimension Class



Base Class

