

DATA WAREHOUSING V.2020

Héctor Álvarez Gómez
Felipe Vildoso Castillo



Ingeniería Industrial

FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE



Ingeniería Industrial
FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

MODELOS DE DATOS

Capítulo 2



Ingeniería Industrial
FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

2.4 MODELOS DE DATOS MULTIDIMENSIONALES

Capítulo 2

DEFINICIONES

Bases de Datos Multidimensionales

- También se llama MDDB.
- Se trata de un producto de software que puede almacenar y administrar datos **usando el enfoque multidimensional**.

Herramientas OLAP

- Conjunto de productos de software diseñado para **facilitar el análisis multidimensional**.
- Puede incorporar adquisición de datos, acceso a los datos, manipulación de datos, o cualquier otra combinación de los conceptos anteriores.

MODELAMIENTO RELACIONAL

Las bases de datos relacionales están organizadas en **torno de una lista de registros**.

Cada registro contiene información que es organizada en campos.

Un ejemplo típico podría ser una lista de clientes con campos como la dirección y el teléfono.

Nombre	Id	Teléfono	Dirección
Banco de Chile	20454	565-8748	Alameda 1200
Endesa	24545	326-4569	Sta. Rosa 55
CTC	17365	215-4563	Providencia 11

MODELAMIENTO RELACIONAL

A pesar de que la tabla tiene varias columnas de información, **cada segmento se relaciona solamente a un nombre de cliente.**

Si se trata de crear una matriz bi-dimensional con el nombre del cliente, por un lado, y cualquier otro campo por el otro (como el teléfono), se podría ver que la correspondencia es uno a uno.

Esta tabla **no debe prestarse** para ser utilizada en una base de datos multidimensional.

Se puede apreciar que cuando se tienen **datos de ventas de productos en cada región** en una tabla relacional hay una correspondencia entre los campos mayor que uno a uno.

MODELAMIENTO RELACIONAL

Producto	Región	Venta
Leche	Norte	1.000.000
Leche	Centro	4.500.330
Leche	Sur	500.867
Queso	Norte	3.567.092
Queso	Centro	2.456.752
Queso	Sur	1.345.698
Crema	Norte	900.345
Crema	Centro	725.938
Crema	Sur	1.987.000

MODELAMIENTO MULTIDIMENSIONAL

Una manera más clara para representar la tabla anterior puede ser la siguiente:

	Norte	Centro	Sur
Leche	1.000.000	4.500.330	500.867
Queso	3.567.092	2.456.752	1.345.698
Crema	900.345	725.938	1.987.000

MODELAMIENTO MULTIDIMENSIONAL

Los datos de ventas corresponden a una matriz de 2 dimensiones: Producto y Región.

A pesar de que los datos pueden ser almacenados en una tabla relacional de 3 campos, **es mucho más natural hacerlo en una matriz de 2 dimensiones.**

En lenguaje multidimensional podríamos decir que esta matriz representa las Ventas dimensionadas por Producto y Región.

MODELAMIENTO MULTIDIMENSIONAL

Si las consultas que se requiere responder son del tipo ¿Cuál fue la venta de Leche en el norte? o ¿Cuál fue la venta de Queso en el sur? y otras consultas que retornan sólo un valor, entonces **no es necesario poner los datos en una base de datos multidimensional**.

Sin embargo, si se quiere responder a preguntas como ¿Cuáles fueron las ventas totales de leche? o ¿Cuáles fueron las ventas totales en el norte? entonces estamos hablando de consultas que **involucran la recuperación de múltiples valores y su agregación**.

El tiempo de respuesta de una base de datos multidimensional, sin embargo, depende de la **cantidad de valores que han sido pre-totalizados**.

MODELAMIENTO MULTIDIMENSIONAL

Si queremos que el tiempo de respuesta sea consistentemente rápido independiente de la consulta realizada, entonces es **necesario mantener consolidados todos los totales y subtotales lógicos**.

En el ejemplo se tiene que la matriz bi-dimensional queda de la siguiente forma:

	Norte	Centro	Sur	Total
Leche	1.000.000	4.500.330	500.867	6.001.197
Queso	3.567.092	2.456.752	1.345.698	7.369.542
Crema	900.345	725.938	1.987.000	3.613.283
Total	5.467.437	7.683.020	3.833.565	16.984.022

MODELAMIENTO MULTIDIMENSIONAL

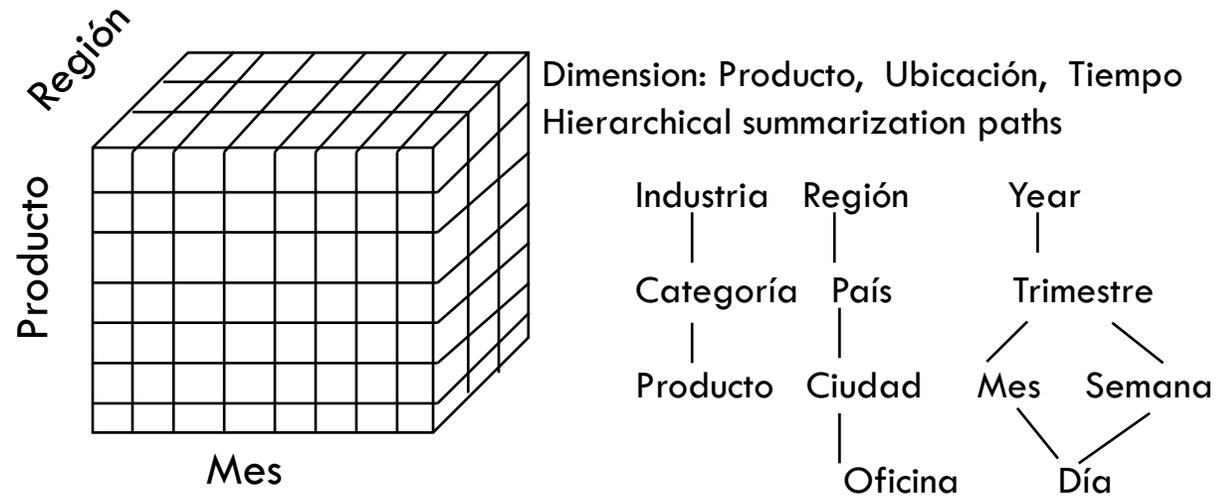
Jerarquías. Los miembros de una dimensión **pueden ser organizados en relaciones de padre-hijo**, donde típicamente el miembro padre representa la consolidación de los miembros que tiene por hijos. Esta organización es llamada jerarquía, y las relaciones padre-hijo son llamadas relaciones jerárquicas.

Una propiedad importante de **las jerarquías es que pueden colapsar y expandir**, es decir, navegar de un nivel de datos detallado a uno más sumariado o resumido, y viceversa.

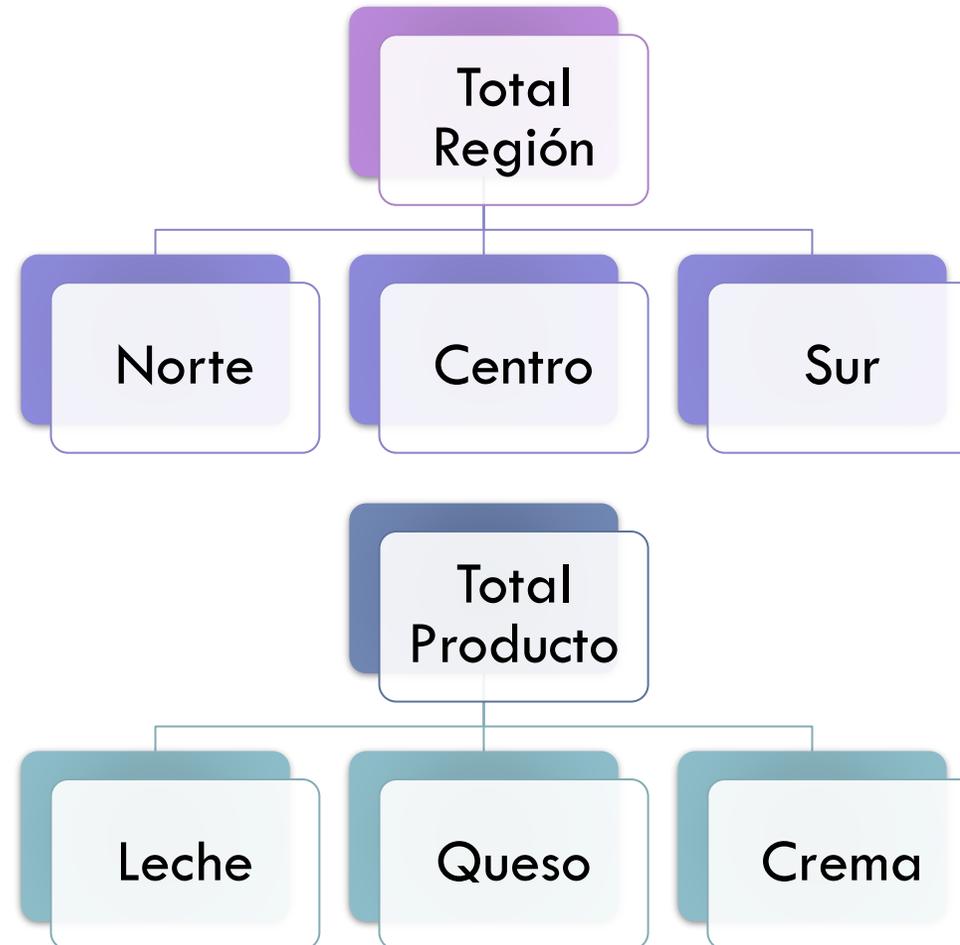
Existen dos tipos de jerarquías, las simples y las compuestas. Las simples son aquellas en que los diferentes niveles colapsan en un solo padre. Las compuestas son aquellas en que un nivel puede colapsar en dos o más padres distintos. En el ejemplo anterior, podemos ver que existen dos jerarquías simples, una en Región y otra en Producto.

EL CONCEPTO DE JERARQUIA

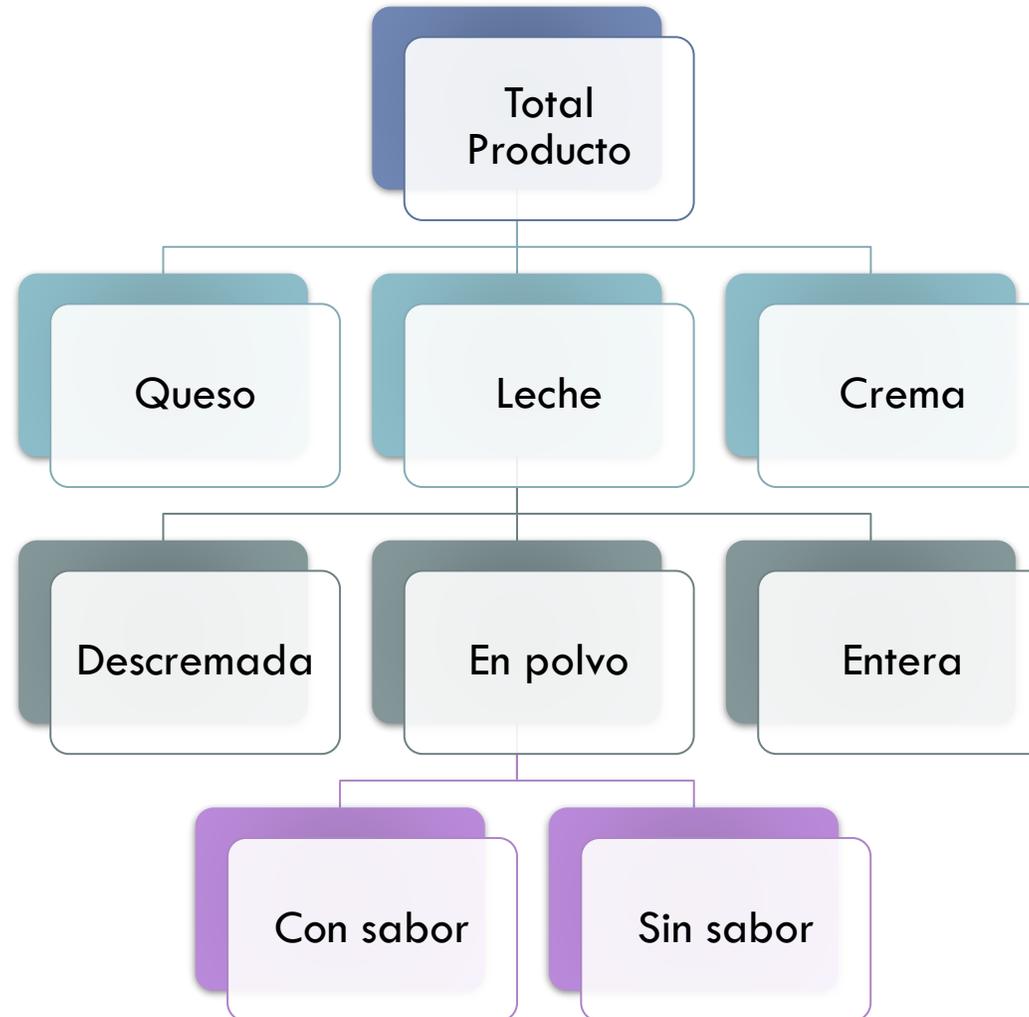
El concepto de jerarquía permite que los datos sean manejados en varios niveles de abstracción.



MODELAMIENTO MULTIDIMENSIONAL

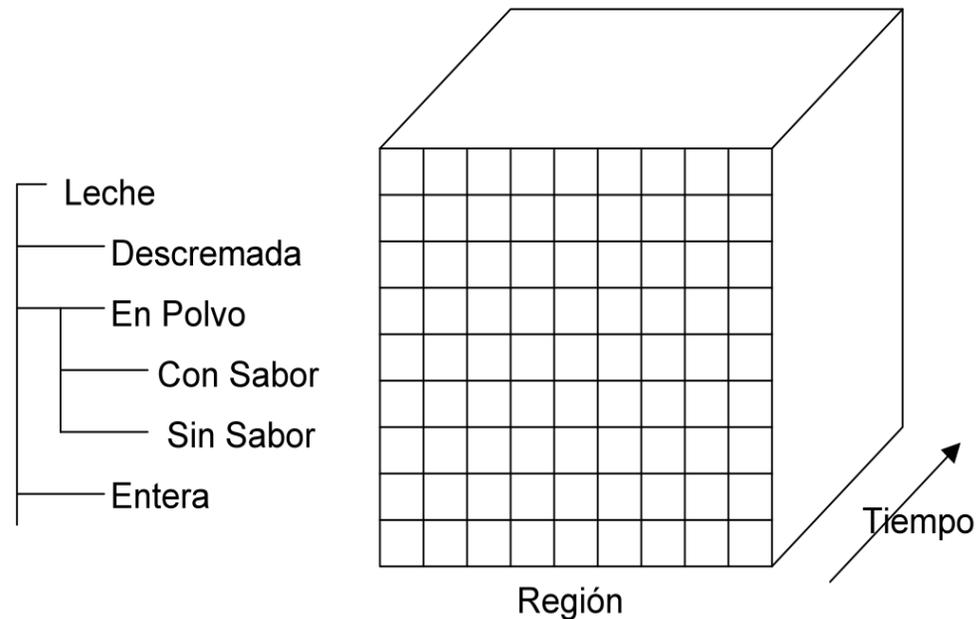


MODELAMIENTO MULTIDIMENSIONAL



MODELAMIENTO MULTIDIMENSIONAL

Ahora podríamos introducir una tercera dimensión, por ejemplo el tiempo. Por lo tanto, la estructura para la base de datos multidimensional es finalmente la siguiente:



MODELAMIENTO MULTIDIMENSIONAL

Podríamos tener las ventas por canal (Retail, Directa, Especial), en el tiempo (anual, mensual, etc), y por tipo de producto (Audio, video, etc). También podríamos agregar una nueva dimensión, como la localización geográfica (Asia, Europa, América).

	Retail	Directo	Especial
Audio	300	350	100
Asia	100	50	20
Europa	100	100	30
América	100	200	50
Video	150	300	400
Asia	30	50	100
Europa	50	100	200
América	70	150	100

MODELAMIENTO MULTIDIMENSIONAL

Los recientes avances tecnológicos han permitido administrar datos como nunca antes se había podido en la historia de la computación. Lo que ha hecho que los grandes **fabricantes de software han agregado a sus productos funcionalidades OLAP.**

Los RDBMS siguen siendo la mejor tecnología que existe para administrar datos a un nivel atómico, pero la **tecnología que implementa los MDDB está mejorando.**

EL modelamiento estrella es una buena alternativa para crear ambientes multidimensionales en bases de datos relacionales. Sin embargo, no se recomienda para almacenar datos con un gran nivel de desagregación (nivel atómico).



Ingeniería Industrial
FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

2.4 OLAP, ROLAP Y MOLAP

Capítulo 2

¿QUE ES OLAP?

ONLINE ANALYTICAL PROCESSING

Es un conjunto de funcionalidades desarrolladas para facilitar el análisis multidimensional.

OLAP es multidimensional porque para responder una consulta compleja, muchas veces recurre a la combinación de varias fuentes de datos.

Interfaz
de
Usuario

- Formato de planilla electrónica.
- Despliega datos desde diferentes perspectivas.
- Realiza análisis estadísticos y test sobre los datos.
- Consulta los datos a un alto nivel de detalle.
- Visualiza los datos con herramientas gráficas y tablas.

ROLAP V/S MOLAP

MOLAP

- Multidimensional OLAP.
- Usan MDDBS para almacenar y acceder a los datos.
- MDDBS “directamente” maneja los datos multidimensionales.
- Usualmente requiere de herramientas de acceso propietarias (no usa SQL).

ROLAP

- Relational OLAP.
- Usan RDBMS para implementar un ambiente OLAP.
- Tipicamente involucra un esquema estrella que provee capacidades multidimensionales.

ROLAP V/S MOLAP: ALMACENAMIENTO

ROLAP	MOLAP
Las sumalizaciones son efectuadas por el RDBMS	Las sumalizaciones son efectuadas por el MMD
Los datos se almacenan en tablas de la estrella	Datos almacenados en arreglos
Metadatos se almacenan en tablas relacionales	En la celda, se almacena el dato actual.
	Los índices sólo se preocupan para distribuir la administración de los datos y apoyar su uso en RAM.

ROLAP V/S MOLAP: MANIPULACIÓN

ROLAP	MOLAP
Hay funciones, por ejemplo ranking, promedio, móviles, etc, que no son soportadas por las funciones clásicas de SQL.	Gran capacidad de administración de muchas dimensiones.
Parte del cálculo se realiza fuera del motor (depende de la herramienta que se use) .	Labores de rotación, cortes, etc. Son parte del motor.

ROLAP V/S MOLAP: RENDIMIENTO

ROLAP	MOLAP
Posee un gran rendimiento, pero se requiere una mantención constante de las estructuras de acceso rápido.	Mantienen pre calculado casi todo, por lo que el acceso al dato es muy rápido.
Requiere de tablas de sumalización para acelerar ciertas consultas.	El proceso de carga es más lento que en un motor relacional (que requiere preprocesar los archivos, principalmente ordenarlos).

OJO: Hay más productos y motores ROLAP que MOLAP.



Ingeniería Industrial
FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

2.5 OPERACIONES QUE SE EFECTUARÁN SOBRE LOS DATOS

Capítulo 2

OPERACIONES OLAP

Como hemos visto, lo importante es disponer de consultas rápidas. Más específicamente es necesario saber cuáles son los **tipos de consultas** que se pueden realizar de forma de ajustar el diseño.



SLICE

Corresponde a la operación si fijamos una dimensión, lo que reduce la dimensionalidad del modelo.

Por ejemplo, cuando Año es igual a 2020.

Comuna	Año	Ventas
Santiago	2020	100
La Cisterna	2020	100

DICE

Corresponde a un sub cubo, usando condiciones en dos o más dimensiones.

Por ejemplo, cuando Año es igual a 2020 o Comuna es igual Iquique.

Año	Comuna	Ventas
2020	Santiago	100
2019	Iquique	563

ROLL-UP O DRILL-UP

Corresponde a la operación que nos permite ver la data sumariada en alguna de sus jerarquías.

Comuna	Ventas
Santiago	100
La Cisterna	100
La Serena	250
Iquique	563



Región	Ventas
Metropolitana	200
Coquimbo	250
Tarapacá	563

DRILL-DOWN

Contrario al Drill-Up, corresponde a la operación que nos permite ver la data con mayor detalle.

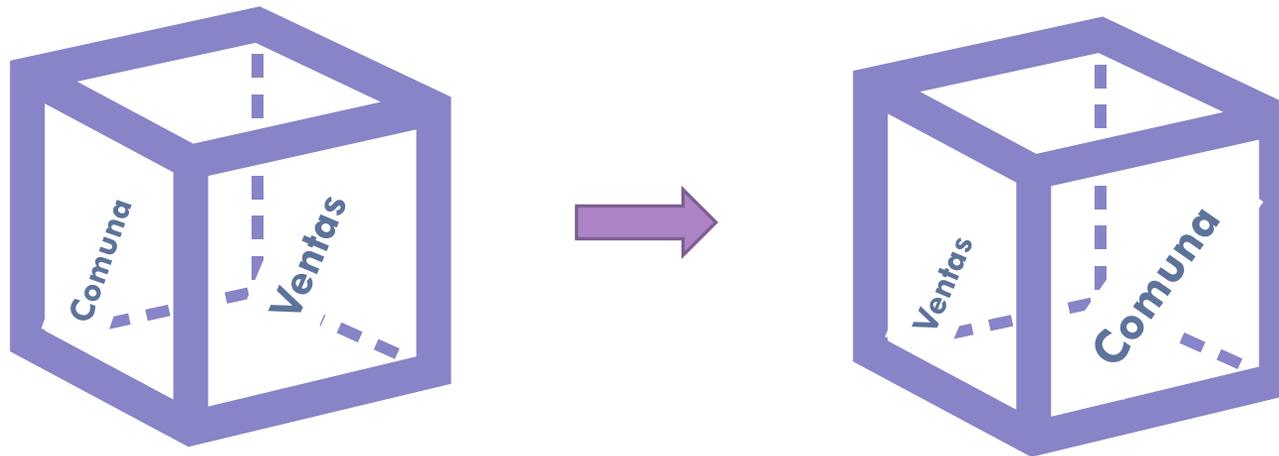
Región	Ventas
Metropolitana	200
Coquimbo	250
Tarapacá	563



Comuna	Ventas
Santiago	100
La Cisterna	100
La Serena	250
Iquique	563

PIVOT O ROTAR

Corresponde a invertir los ejes para tener distintas perspectivas.





fcfm

Ingeniería Industrial

FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE



DATA WAREHOUSING V.2020

Héctor Álvarez Gómez
Felipe Vildoso Castillo