



Grupo #1

Es una base de datos corporativa en la que se integra información depurada de las diversas fuentes que hay en la organización.



Dimensiones y Hechos

Los datos relacionales, aquellos manejados por los gestores de base de datos convencionales, pueden ser considerados bidimensionales, ya que cada unidad de datos, denominada hechos, está correlacionada con una fila y una columna, cada una de las cuales puede ser considerada una dimensión.

Dimensiones

Representan factores por lo que se analiza un determinado área del negocio. Son pequeñas y usualmente están desnormalizadas.

Hechos

Son el objeto de los análisis y están relacionados con las dimensiones. Son tablas muy grandes y suelen estar desnormalizadas. Se a menudo incluyen diferentes agregaciones como máximo, mínimo, media.

Los datos relacionales, aquellos manejados por los gestores de bases de datos convencionales, pueden ser considerados bidimensionales, ya que cada unidad de dato, denominada hecho, está correlacionada con una fila y una columna, cada una de las cuales puede ser considerada una dimensión.

Una dimensión es una clasificación de alguna actividad en una organización por la cual se puede medir su éxito.

Hecho es el dato en sí mismo, los miembros son el conjunto de datos que describen una dimensión.

Funcionalidades y beneficios

En un Data Warehouse lo que se quiere es contener datos que son necesarios o útiles para una organización, es decir, que se utiliza como un repositorio de datos para posteriormente transformarlos en información útil para el usuario. Un Data Warehouse debe entregar la información correcta a la gente indicada en el momento óptimo y en el formato adecuado.



Funcionalidades

El Data Warehouse da respuesta a las necesidades de usuarios expertos, utilizando Sistemas de Soporte a Decisiones (DSS), Sistemas de información ejecutiva (EIS) o herramientas para hacer consultas o informes. Los usuarios finales pueden hacer fácilmente consultas sobre sus almacenes de datos sin tocar o afectar la operación del sistema.



En el funcionamiento de un Data Warehouse son muy importantes las siguientes ideas:

Integración de los datos provenientes de bases de datos distribuidas por las diferentes unidades de la organización y que con frecuencia tendrán diferentes estructuras (fuentes heterogéneas).

Separación de los datos usados en operaciones diarias de los datos usados en el almacén de datos para los propósitos de divulgación, de ayuda en la toma de decisiones, para el análisis y para operaciones de control.

- Proporciona información clave para la toma de decisiones empresariales.
- Mejora la calidad de las decisiones tomadas.
- Especialmente útil para el medio y largo plazo.
- Son sistemas relativamente sencillos de instalar si las fuentes de datos y los objetivos están claros.
- Muy útiles para el almacenamiento de análisis y consultas de históricos

- Proporciona un gran poder de procesamiento de información.
- Permite una mayor flexibilidad y rapidez en el acceso a la información.
- Facilita la toma de decisiones en los negocios.
- Las empresas obtienen un aumento de la productividad.
- Proporciona una comunicación fiable entre todos los departamentos de la empresa.

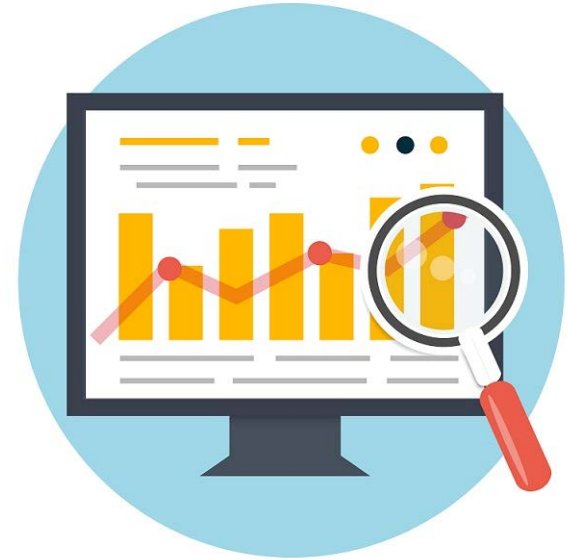
- Mejora las relaciones con los proveedores y los clientes.
- Permite conocer qué está pasando en el negocio, es decir, estar siempre enterado de los buenos y malos resultados.
- Transforma los datos en información y la información en conocimiento
- Permite hacer planes de forma más efectiva.
- Reduce los tiempos de respuesta y los costes de operación.

Herramientas de Consulta e Informes

Las Herramientas de Consulta e Informes son aquellas que facilitan el proceso de consultas de grandes volúmenes de datos almacenados en un Data Warehouse y presentarlas adecuadamente.

Se pueden clasificar en:

1. De Consultas Gestionadas.
2. De Reporte.
 - a. De Reportes de Producción.
 - b. De Redacción de Informes.



Herramientas de Consulta e Informes



Las Herramientas de Consulta Gestionadas se encargan de esconder la complejidad de las estructuras de base de datos por medio de una **interfaz/capa o meta-layer**.

Esta capa provee al usuario de vistas **orientadas a temas y (generalmente) son amigables** con un sistema “point and click”. Presentan la información como **informes fáciles de leer o en dashboards**.

Este tipo de herramientas son la mejor opción para usuarios finales de aplicaciones de negocios.

Las Herramientas de Reportes de Producción permiten generar informes de las operaciones regulares o de apoyo de gran volúmen de la empresa (como cálculo e impresión de cheques).

Los Redactores de Informes, por el contrario, son herramientas de escritorio simples para usuarios no técnicos, que facilitan la elaboración de reportes.



Fases del Proceso de Consulta:

1. Definición de la consulta.
2. Acceso a Datos Almacenados.
3. Manejo y Aplicación de Cálculos.
4. Elaboración de Reporte.
5. Liberación de Reporte.



Tipos y Planificación de las Consultas

Consultas Empaquetadas

Una consulta empaquetada o subconsulta es una consulta (sentencia SELECT) dentro de otra. Por lo general aparece como parte de una condición en las cláusulas WHERE o HAVING. También pueden utilizarse en la cláusula FROM.

Llegan a utilizarse como un procedimiento (consulta empaquetada tipo I) en el que la consulta se ejecuta una vez, o como un ciclo (consulta empaquetada tipo II) en el que la consulta se ejecuta varias veces.

Consultas empaquetadas tipo I

Son como los procedimientos en un lenguaje de programación. Estas consultas se evalúan una vez y se produce una tabla.

Solo se deben usar cuando el resultado no contenga ninguna columna de las tablas en la consulta empaquetada y también para algunos problemas de enlace y de diferencia.

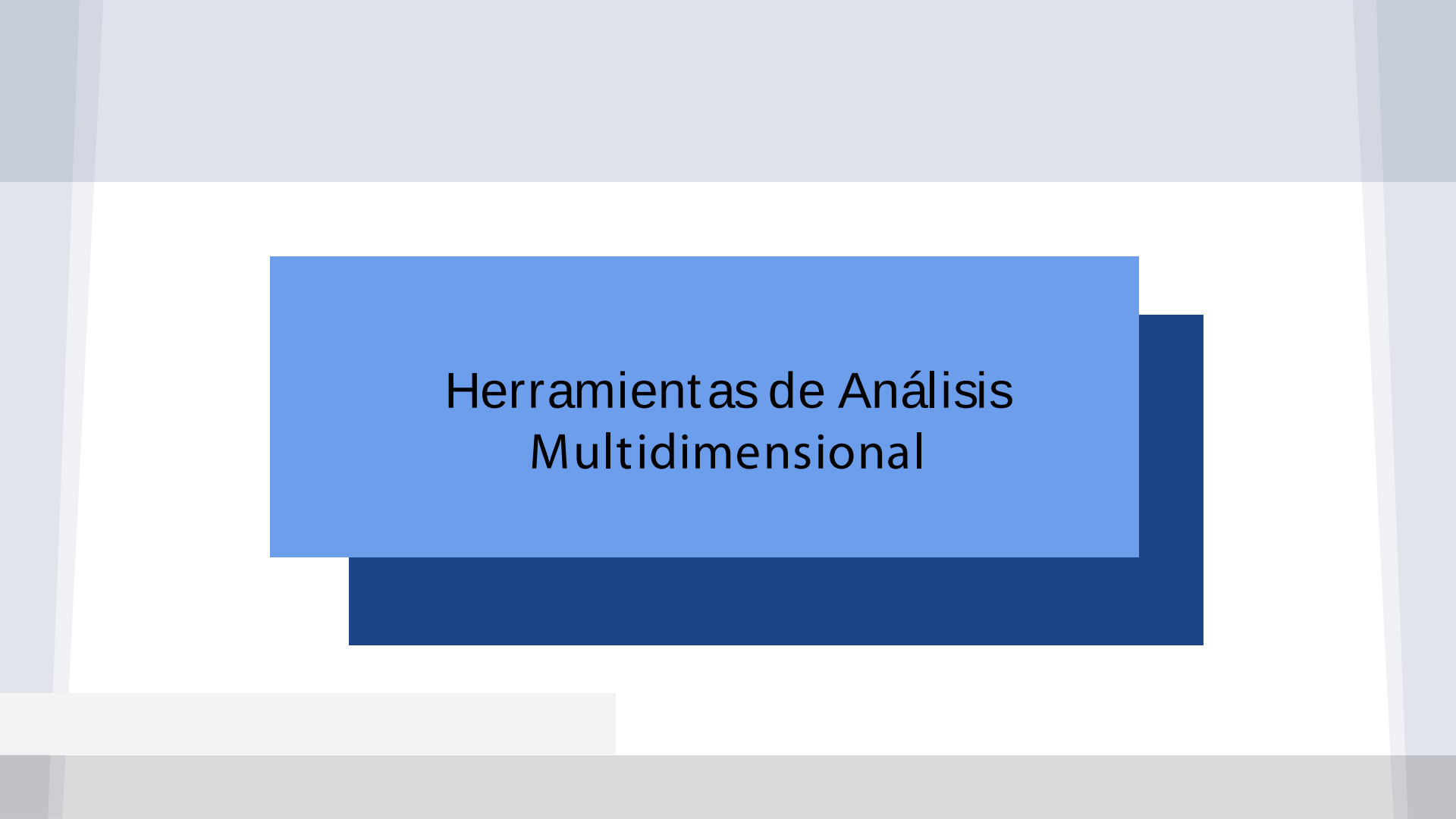
Consultas empaquetadas tipo II

Tienen dos características distintivas; en primer lugar, hacen referencia a una o más columnas desde una consulta externa (subconsultas relacionadas) y en segundo lugar, se ejecutan una vez para cada fila de la consulta externa.

La expresión “*ad-hoc*” identifica que se realiza para un fin determinado; por ende una consulta ad-hoc se realiza específicamente para una determinada ocasión o situación.

Las consultas en base de datos *ad-hoc querying* o *ad-hoc reporting* permiten al usuario personalizar una consulta en tiempo real enves de estar atado a las consultas prediseñadas para informes.

Permiten también a los usuarios con poca experiencia en base de datos, tener el mismo acceso a la información.



Herramientas de Análisis Multidimensional

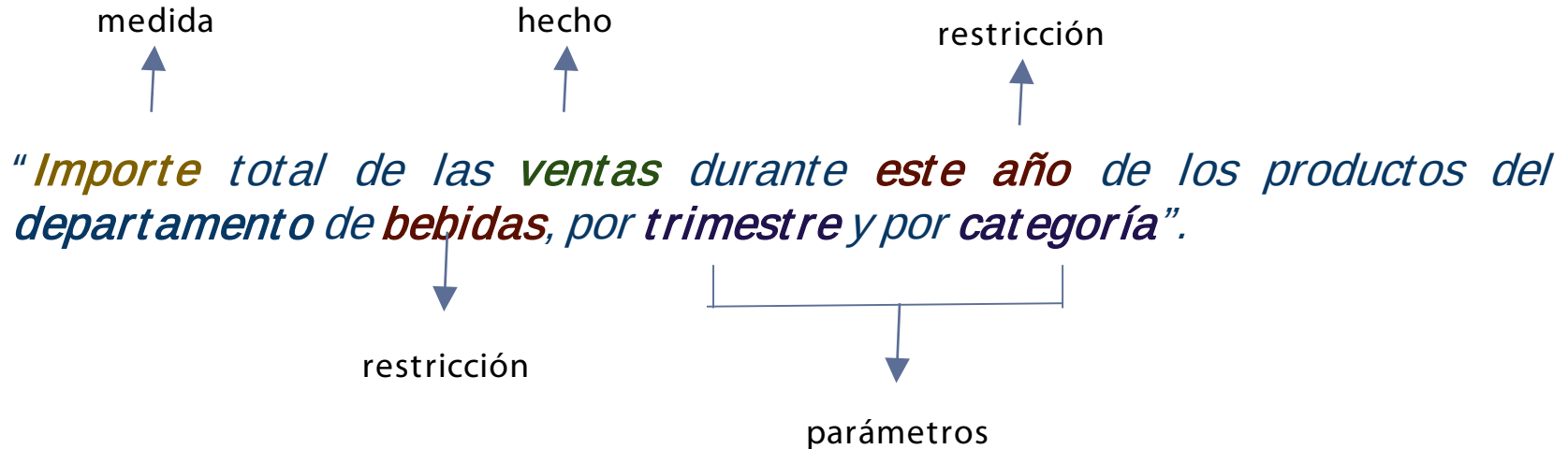
OLAP (*On Line Analytical Processing*)

Análisis multidimensional y dinámico de los datos consolidados de la empresa; ayuda al usuario a sintetizar la información de la empresa a través de vistas personalizadas, análisis histórico y pronósticos.

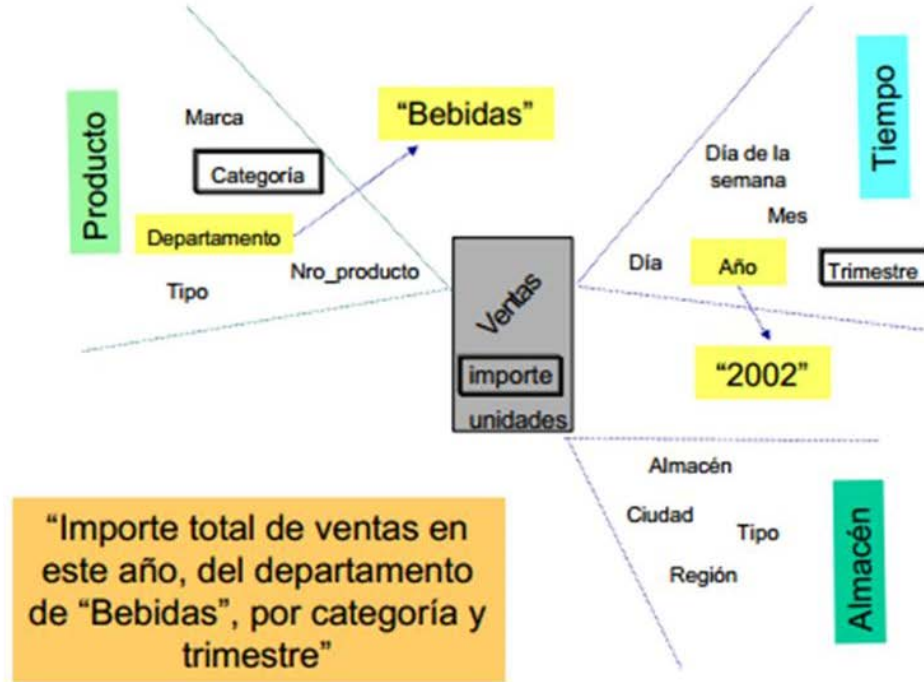
- Está enfocado al análisis de **grandes cantidades de datos**
- Proporciona **respuestas rápidas y complejas**
- Ofrece una **visión multidimensional de los datos** (matrices)
- **No impone restricciones** sobre el número de dimensiones
- Ofrece **simetría para las dimensiones**
- Permite **definir de forma flexible** (sin limitaciones) sobre las dimensiones: restricciones, agregaciones y jerarquías entre ellas.
- Ofrece **operadores intuitivos** de manipulación: *drill-down, roll-up, slice and dice, pivot*

OLAP (On Line Analytical Processing)

Ejemplo:



OLAP (On Line Analytical Processing)



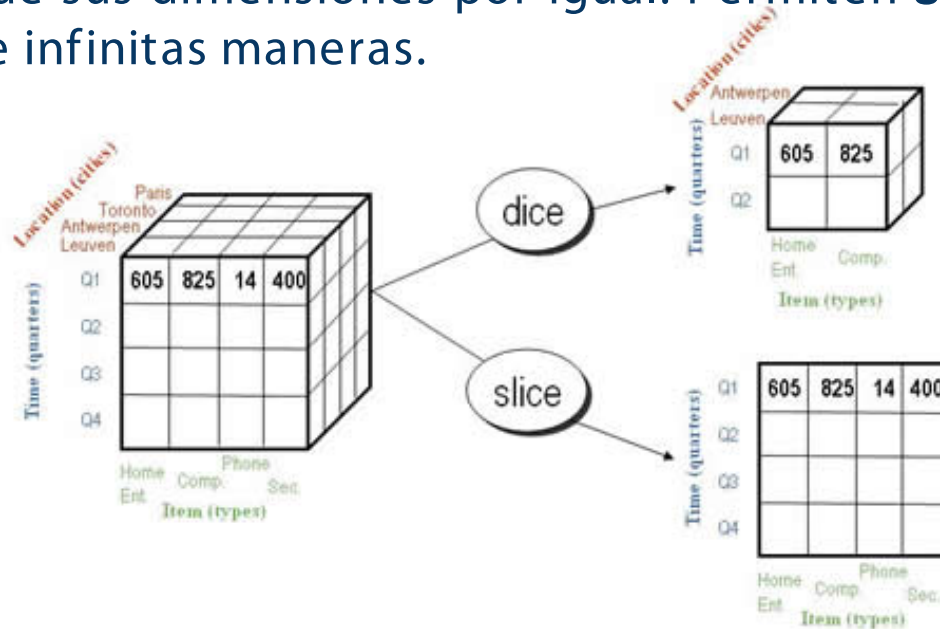
El carácter agregado de las consultas de Análisis de Datos, aconseja la definición de operadores que faciliten el manejo y la presentación de los datos.

Las operaciones conocidas como "*drill down*" y "*roll up*" básicamente van a definir el nivel de granularidad con la que queremos analizar los datos, mientras que las operaciones "*slice*" y "*dice*" nos permiten navegar entre las dimensiones.

Slice & dice, pivot, drill-down, roll-up

Slice & Dice

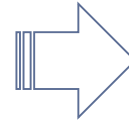
Otorga la habilidad de acceder a un data warehouse a través de cualquiera de sus dimensiones por igual. Permiten **separar y combinar** los datos de infinitas maneras.



Slice & dice, pivot, drill-down, roll-up

Slice & Dice:

	Productos	Tienda 1	Tienda 2
Q1	Juguetes	\$ 5.2	\$ 5.6
	Electrónicos	\$ 1.9	\$ 1.4
	Cosméticos	\$ 2.3	\$ 2.6
Q2	Juguetes	\$ 1.1	\$ 1.1
	Electrónicos	\$ 8.9	\$ 7.2
	Ropa	\$ 0.75	\$ 0.4



	Productos	Tienda 1
Q1	Juguetes	\$ 5.2
	Electrónicos	\$ 1.9
Q2	Juguetes	\$ 1.1
	Electrónicos	\$ 8.9

Pivot

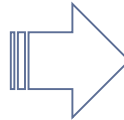
Conocidas como tablas dinámicas, se utilizan en excel para comparar grandes cantidades de datos e intercambiar fácilmente columnas por filas dentro de la misma tabla y realizar filtros que resulten en reportes que de una u otra manera necesitamos en un tiempo considerable para construirlo.

Son una gran herramienta que nos ayuda a realizar un análisis profundo de nuestros datos, ya que podemos filtrar, ordenar y agrupar la información de la tabla dinámica de acuerdo a las necesidades.

Slice & dice, pivot, drill-down, roll-up

Pivot:

	Productos	1	2
Q1	Muebles	\$ 5.2	\$ 5.6
	Juguetes	\$ 1.9	\$ 1.4
	Ropa	\$ 2.3	\$ 2.6
Q2	Muebles	\$ 8.9	\$ 7.2
	Juguetes	\$ 0.7	\$ 0.4
	Ropa	\$ 4.6	\$ 4.6



	Productos	Q1	Q2
1	Muebles	\$ 5.2	\$ 8.9
	Juguetes	\$ 1.9	\$ 0.7
	Ropa	\$ 2.3	\$ 4.6
2	Muebles	\$ 5.6	\$ 7.2
	Juguetes	\$ 1.4	\$ 0.4
	Ropa	\$ 2.6	\$ 4.6

Drill-down y Roll-up

Son utilizadas para mover la vista hacia y desde un mayor nivel de detalle a un menor nivel de detalle.

Para cada caso, hacia un mayor nivel de detalle, realizamos un “*drill down*”, mientras que para un menor nivel de detalle realizamos un “*roll up*”.

Recordando que la granularidad de la información es directamente proporcional al nivel de detalle, por lo que estas operaciones básicamente nos ayudan a ver o analizar los datos desde diferentes niveles de granularidad.

A las operaciones “*drill down*” también se les conoce como agregación dinámica.

Drill-down

Categoría	Trimestre	Ventas
Refrescos	T1	200,000
Refrescos	T2	100,000
Refrescos	T3	300,000
Jugos	T1	150,000
Jugos	T2	240,000
Jugos	T3	100,000



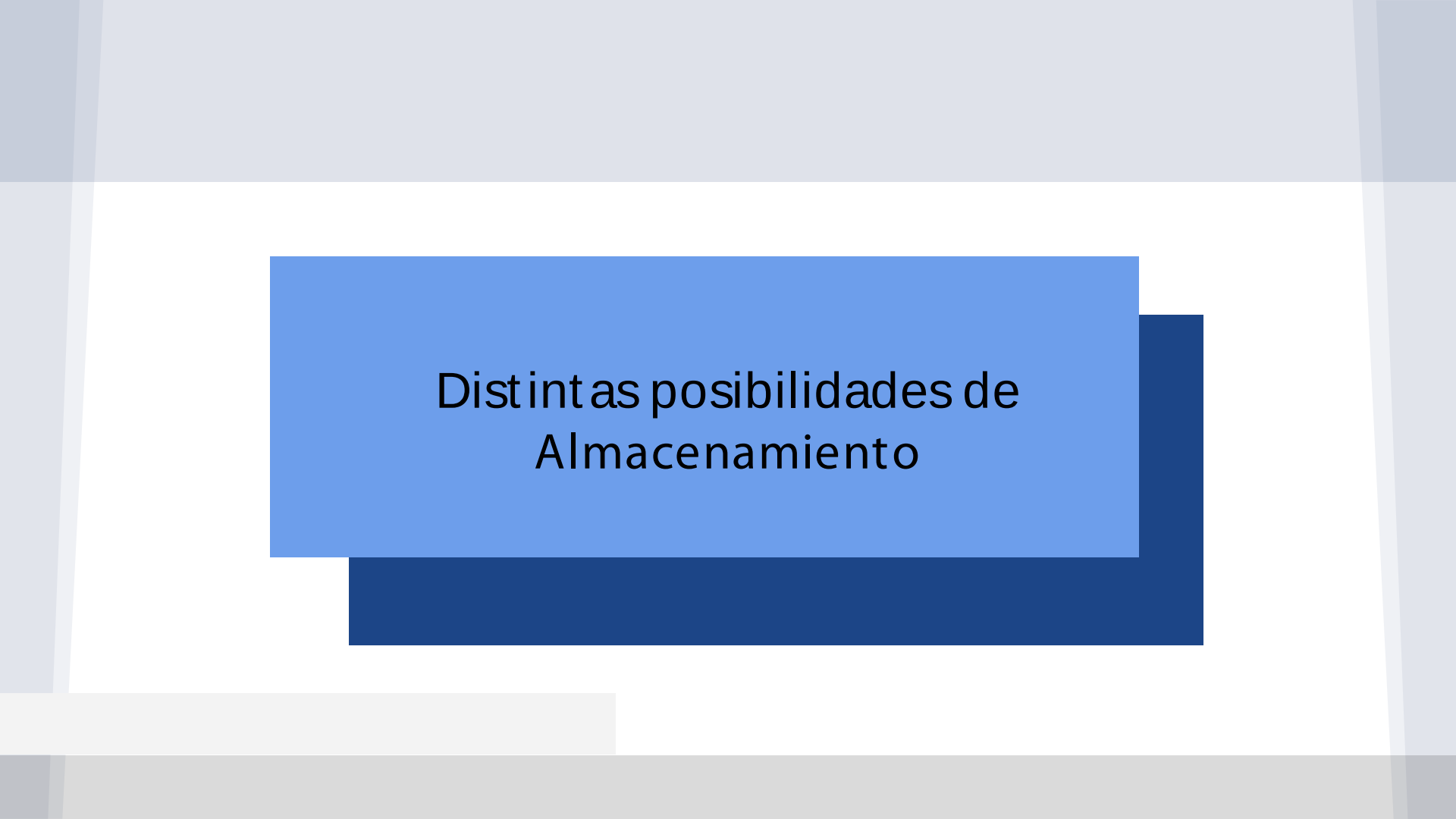
Categoría	Trimestre	Mes	Ventas
Refrescos	T1	Enero	100,000
Refrescos	T1	Febrero	50,000
Refrescos	T1	Marzo	50,000

Roll-up

Categoría	Enero	Febrero	Marzo
Refrescos	100,000	50,000	50,000
Refrescos	25,000	25,000	50,000
Refrescos	100,000	125,000	75,000



Categoría	Trimestre	Ventas
Refrescos	T1	200,000
Refrescos	T2	100,000
Refrescos	T3	300,000



Distintas posibilidades de Almacenamiento

Relacional: ROLAP

OLAP Relacional es lo que todos los datos, incluyendo las agregaciones, se almacenan dentro de una estructura relacional de base de datos que puede estar en la misma localización de la fuente o no.

ROLAP es el método de almacenamiento más lento en la recuperación de datos.

Multidimensional: MOLAP

OLAP Multidimensional, aquí los datos de fuente son almacenados en un formato multidimensional.

Es la opción más rápida para la recuperación de datos, pero requiere de mucho espacio en el disco.

Bibliografía

Tipos y Planificación de las Consultas

*Administración de Base de Datos diseño y desarrollo de Aplicaciones, Michael V. Mannino.
Mc Graw Hill 3er Edición*

Herramientas de Análisis Multidimensional

Principles of Data Mining, Heikki Mannila, David J. Hand and Padhraic Smyth