



Universidad Autónoma Del Estado De México

Unidad de aprendizaje: Data Warehouse I

Unidad 3. Aspectos técnicos en el proceso de creación y explotación del Data Warehouse

Dra. Carmen Liliana Rodríguez Páez



Universidad Autónoma Del Estado De México

DIRECTORIO INSTITUCIONAL

RECTOR

Dr. en Edu. Alfredo Barrera Baca

M. en E.U. y R.

Marco Antonio Luma Pichardo

Secretario de Docencia

M. en C

Jannet Valero Vilchis

Secretaria de Rectoría

Dra. en Ed.

Sandra Chávez Marín

Secretaria de Extensión y Vinculación

M. en Dis.

Juan Miguel Reyes Viurquez

Secretario de Administración

M. en L.A.

María del Pilar Ampudia García

Secretaria de Cooperación Internacional

Lic. en Com.

Gastón Pedraza Muñoz

Director General de Comunicación Universitaria

M. en D.F.

Jorge Rogelio Zenteno Domínguez

Encargado del Despacho de la Contraloría Universitaria

M. en A.

José Francisco Mejía Carbajal

Secretario Particular Adjunto del Rector

Dr. en C.I.

Carlos Eduardo Barrera Díaz

Secretario de Investigación y Estudios Avanzados

Dr. en A.

José Édgar Miranda Ortiz

Secretario de Difusión Cultural

M. en E.

Javier González Martínez

Secretario de Finanzas

Dr. en C.C.

José Raymundo Marcial Romero

Secretario de Planeación y Desarrollo Institucional

Dra. en Dis.

Mónica Marina Mondragón

Secretaría de Cultura Física y Deporte

M. en A. P.

Guadalupe Ofelia Santamaría González

Directora General de Centros Universitarios y Unidades Académicas

Profesionales

Lic. En Act.

Angelita Garduño Gómez

Secretaria particular del Rector



Universidad Autónoma Del Estado De México

DIRECTORIO DEL CENTRO UNIVERSITARIO UAEM NEZAHUALCOYÓTL

**Maestro en Derecho Juan Carlos Medina Huicochea
ENCARGADO DEL DESPACHO DE LA DIRECCIÓN**

Maestro en Ciencias

José Antonio Castillo Jiménez Subdirector Académico

Licenciado en Economía

Ramón Vital Hernández Subdirector Administrativo

Doctora en Ciencias Sociales

María Luisa Quintero Soto Coordinadora de Investigación y Estudios Avanzados

Licenciado en Administración de Empresas

Víctor Manuel Durán López Coordinador de Planeación y Desarrollo Institucional

Maestro en Ciencias

Cesar Lucio Gutiérrez Ruiz Coordinador de la Licenciatura en Comercio Internacional

Maestro en S.F.

Carlos Anaya Hernández Coordinadora de la Licenciatura en Educación para la Salud

Doctor en Ingeniería de los Sistemas

Ricardo Rico Molina Coordinador de la Licenciatura en Ingeniería en Sistemas Inteligentes

Maestro en Ciencias

Ricardo Pacheco Ruiz Coordinador de la Licenciatura en Ingeniería en Transporte

Maestro en Ciencias de la Computación

Erick Nicolás Cabrera Álvarez Coordinador de la Licenciatura en Seguridad Ciudadana Mixta

Maestro en Administración

José Ramon CS. Garcia Ibarra Coordinador de la Licenciatura en Seguridad Ciudadana Presencial



Universidad Autónoma Del Estado De México

UBICACIÓN DE LA ASIGNATURA DATA WAREHOUSE I

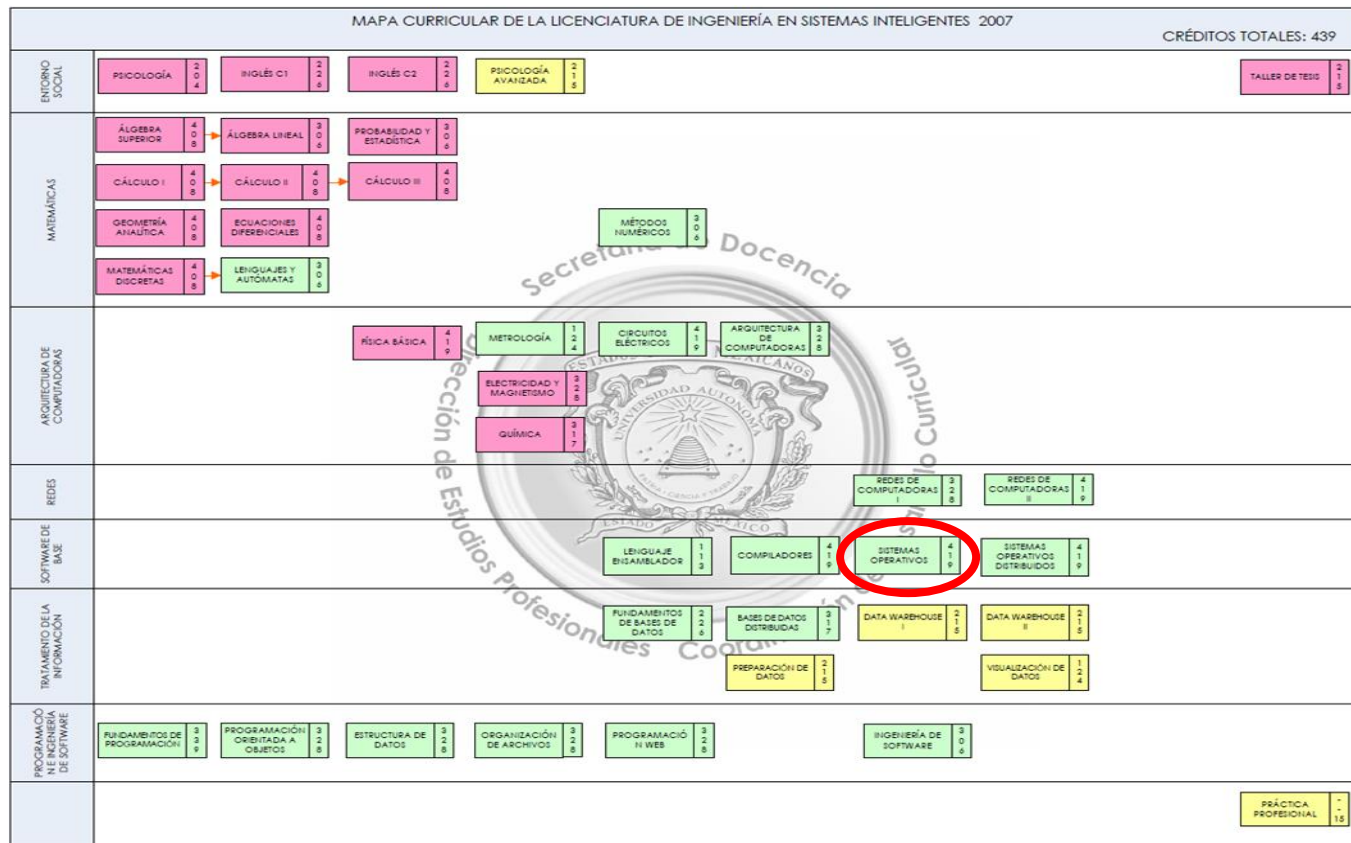


Figura 1. Mapa Curricular ISI-Data Warehouse I



Universidad Autónoma Del Estado De México

UNIDAD DE APRENDIZAJE DATA WAREHOUSE I

La Unidad de Aprendizaje (UA) de *Data Warehouse I* tiene como área curricular el tratamiento de la información y forma parte del núcleo sustantivo.



Universidad Autónoma Del Estado De México

PRESENTACIÓN DEL MATERIAL

El material de visión proyectable de la Unidad de Aprendizaje Data Warehouse I es un documento que contiene los temas: OLTP, OLAP, y MOLAP. Correspondientes a la unidad III. del programa por competencia.



Universidad Autónoma Del Estado De México

Este material permitirá a los alumnos la comprensión de los temas, y será una guía de apoyo para su elaboración del proyecto final.

Así mismo, se busca que adquiriera la habilidad para aplicar los modelos para la creación de programas de cómputo y software que de solución a problemas de almacenamiento, procesamiento y acceso de información.



Universidad Autónoma Del Estado De México

OBJETIVO DE LA UNIDAD DE APRENDIZAJE

Comprender los fundamentos teóricos básicos del Data Warehouse para el análisis, mapeo, y transformación de datos útil en la toma de decisiones.



Universidad Autónoma Del Estado De México

UNIDAD DE COMPETENCIA

Unidad 3. Aspectos técnicos en el proceso de creación y explotación del Data Warehouse

Objetivo: Identificar indicadores y perspectivas para el desarrollo de un Data Warehouse

Contenidos:

3.1. Análisis de Requerimientos de la empresa

3.1.1 Indicadores y perspectivas

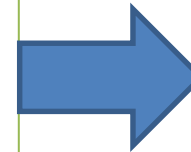
3.2 Modelado de datos

3.2.1 OLAP

3.2.2 OLTP

3.2.3 MOLAP

3.3 Factores relevantes en el proceso de adquisición



3.2.1 OLAP

3.2.2 OLTP

3.2.3 MOLAP



Universidad Autónoma Del Estado De México

¿QUÉ ES OLAP?

OLAP (On-Line Analytical Processing), es un método para organizar y consultar datos sobre una estructura multidimensional proporcionando una mayor agilidad y flexibilidad al usuario de negocio.

Forma parte de lo que se conoce como sistemas analíticos, esta formada por un motor y un visor.



Universidad Autónoma Del Estado De México

Ventajas

- ✓ Tiene acceso a grandes cantidades de información.
- ✓ Analiza las relaciones entre muchos tipos de elementos empresariales.
- ✓ Compara datos agregados a través de periodos jerárquicos.
- ✓ Involucran cálculos complejos entre elementos de datos.
- ✓ Pueden responder con rapidez a consultas de usuarios.

Desventajas

- Imposibilidad de realizar cambios en su estructura.
- Requiere organizar los datos en un esquema estrella el cual es complicado de implementar y administrar.
- No puede tener un gran numero de dimensiones en un único cubo OLAP.



Universidad Autónoma Del Estado De México

Operaciones analíticas básicas de la OLAP

Cuatro tipos de operaciones analíticas en OLAP son:

- Roll-up
- Desglose
- Cortar y dados
- Pivotar (girar

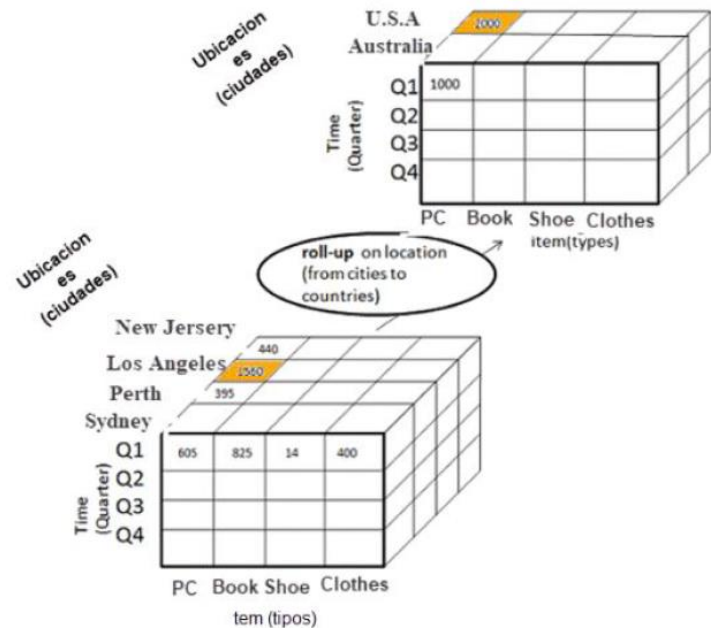


Universidad Autónoma Del Estado De México

Roll-up:

También se conoce como “Consolidación” o “agregación”.

1. Reducir dimensiones
2. Subiendo la jerarquía conceptual. La jerarquía conceptual es un sistema de agrupación de las cosas en función de su orden o nivel.



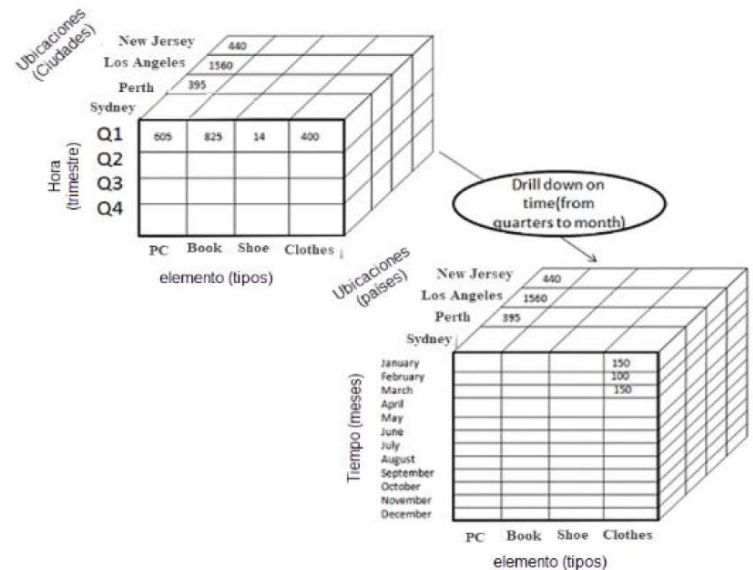


Universidad Autónoma Del Estado De México

Desglose

En el desglose los datos se fragmentan en partes mas pequeñas. Es lo opuesto al proceso acumulativo. Se puede hacer a través de:

- Bajar la jerarquía conceptual
- Aumento de una dimensión.

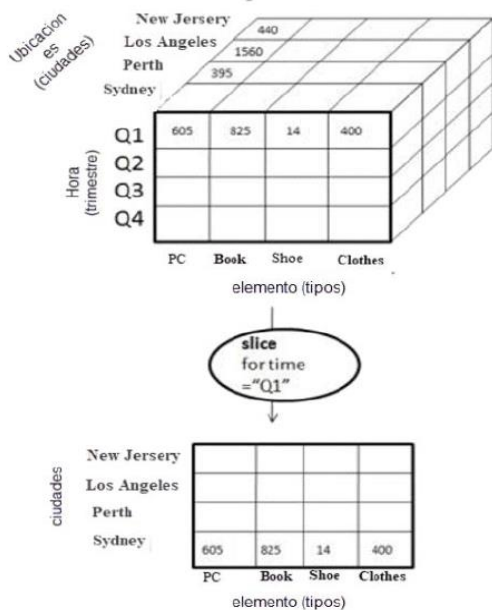




Universidad Autónoma Del Estado De México

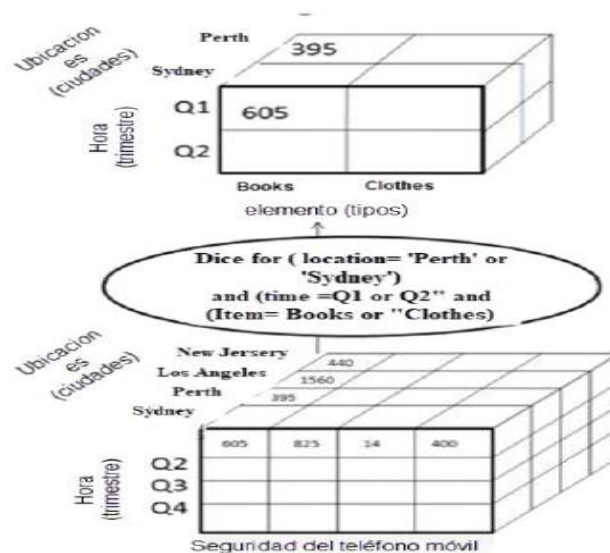
Rebanada

Aquí se selecciona una dimensión y se crea un nuevo subcubo.



Dados

Esta operación es similar a un segmento. La diferencia en datos es que selecciona 2 o mas dimensiones que dan como resultado la creación de un subcubo.





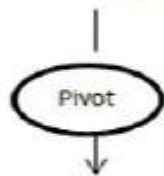
Universidad Autónoma Del Estado De México

Pivot

Ubicación es (ciudades)

New Jersey				
Los Angeles				
Perth				
Sydney	605	825	14	400
	PC	Book	Shoe	Clothes

Artículo (tipos)



Artículo (tipos)

PC				605
Book				825
Shoe				14
Clothes				400
	New Jersey	Los Angeles	Perth	Sydney

Ubicación (Ciudades)

El Pivot girara los ejes de datos para proporcionar una presentación sustituta de los datos.



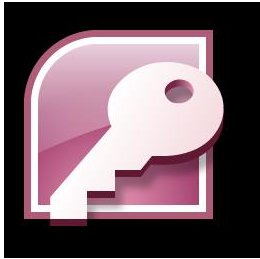
Universidad Autónoma Del Estado De México

Tipos de sistemas OLAP

Sistemas	Ventajas	Desventajas
<p>ROLAP</p> <p>Funciona con datos que existen en una base de datos relacional.</p> <p>Las tablas de hechos y dimensiones se almacenan como tablas relacionales.</p>	<p>Alta eficiencia de datos porque el rendimiento de las consulta y el lenguaje de acceso están optimizados para el análisis de datos multidimensionales.</p> <p>Escalabilidad: Ofrece escalabilidad para administrar grandes volúmenes de datos.</p>	<p>Demanda de recursos mas elevados ya que necesita una alta utilización de mano de obra, software y hardware.</p> <p>Rendimiento de consultas lento en comparación con MOLAP.</p>
<p>HOLAP</p> <p>Es una mezcla de ROLAP y MOLAP ya que ofrece un rápido cálculo y una escalabilidad mayor a la de ROLAP.</p> <p>Utiliza 2 bases de datos</p> <p>1-Datos calculados almacenados en un cubo OLAP</p> <p>2-Información detallada almacenada en una base de datos relacional.</p>	<p>Rendimiento mas rápido debido a que utiliza tecnología de cubo.</p> <p>Acceso a datos actualizados en tiempo real</p> <p>Ahorra espacio en disco y sigue siendo compacto, lo que ayuda a evitar problemas relacionados con la velocidad de acceso.</p>	<p>Mayor nivel de complejidad porque soporta herramientas y aplicaciones ROLAP Y MOLAP.</p> <p>Mayores posibilidades de solapamiento especialmente es sus funcionalidades.</p>



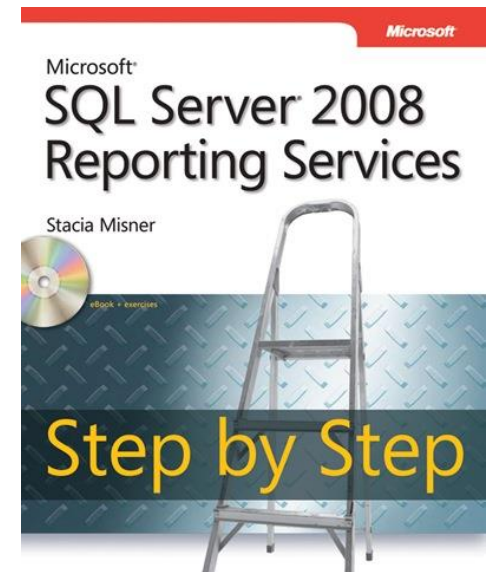
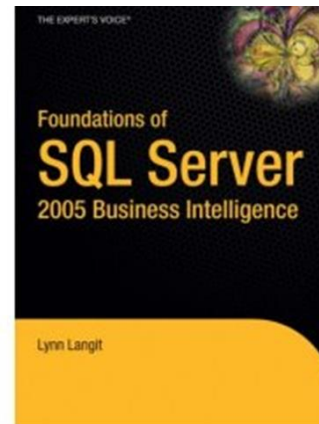
Utilización de Herramientas para OLAP



ORACLE®



Commercial Open Source
Business Intelligence Software





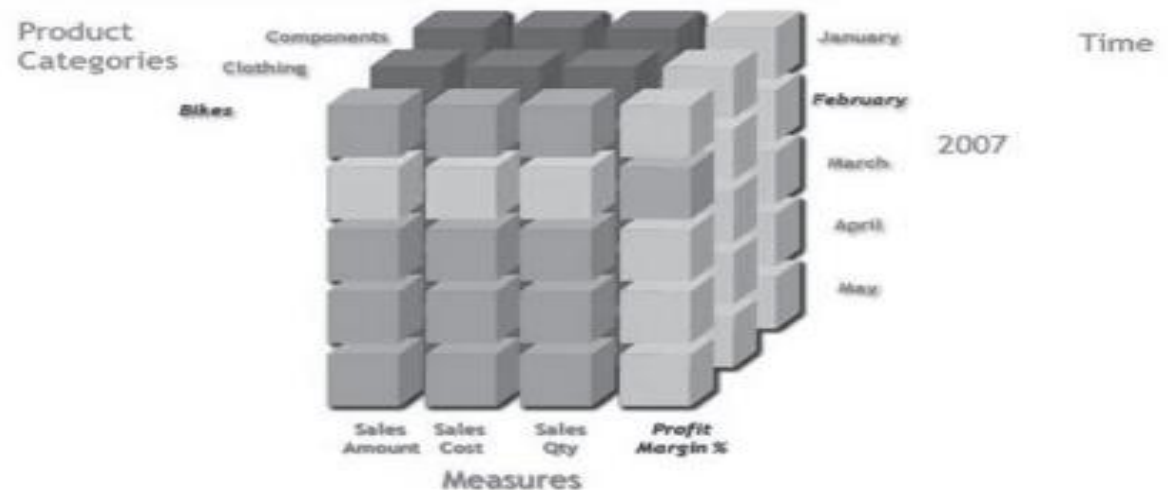
Universidad Autónoma Del Estado De México

Ejemplo

Amazon analiza las compras realizadas por sus clientes para crear una pagina de inicio personalizada con productos que probablemente interesen a sus clientes.

¿Cuál es el margen de beneficios de la venta de bicicletas para febrero de 2007?

- Tiempo
- Productos
- Medidas





Universidad Autónoma Del Estado De México

OLTP

- Es el procesamiento de transacciones en línea, admite aplicaciones orientadas a transacciones en una arquitectura de 3 niveles.
- Administra las transacciones diarias de una organización





Universidad Autónoma Del Estado De México

OLTP

Objetivo Principal

Procesamiento de datos y no el análisis de datos.



Universidad Autónoma Del Estado De México

Beneficios del método

- Administra las transacciones diarias de una organización.
- OLTP amplía la base de clientes de una organización al simplificar los procesos individuales.

Inconvenientes del método

- Si el sistema se enfrenta a fallos de hardware las transacciones en línea se ven gravemente afectadas.
- Los sistemas permiten a varios usuarios acceder y cambiar los mismos datos al mismo tiempo, lo que muchas veces crea una situación sin precedentes.



Universidad Autónoma Del Estado De México

C A R A C T E R Í S T I C A S

Parámetros	Descripción	Parámetros	Descripción
Proceso	Sistema transaccional en línea que gestiona la modificación de las bases de datos	Audiencia	Proceso orientado al mercado
Característico	Gran numero de transacciones en línea cortas	Tipo de consulta	Son estandarizadas y simples
Funcionalidad	Sistema de modificación de bases de datos en línea	Respaldo	Copia de seguridad completa de los datos combinada con backups incrementales
Método	Utiliza DBMS tradicional	Diseño	Esta orientado a la aplicación.
Consulta	Insertar, actualizar y eliminar información de la base de datos	Tipo de usuario	Utilizado por usuarios críticos de datos, como empleados, DBA y profesionales de BD
Cuadro	Las tablas de la BD se normalizan	Propósito	Diseñado para operaciones empresariales en tiempo real
Fuente	El método y sus transacciones son las fuentes de datos	Métrica de rendimiento	Rendimiento se transacciones
Integridad de datos	La BD debe mantener la restricción de la integridad de los datos	Numero de usuarios	Permite miles de usuarios
Tiempo de respuesta	En milisegundos	Productividad	Aumenta el autoservicio y la productividad del usuario
Calidad de los datos	Los datos se detallan y organizan	Proceso	Proporciona un resultado rápido para los datos de uso diario
Utilidad	Ayuda a controlar y ejecutar tareas empresariales fundamentales.	Característica	Es fácil de crear y mantener
Operación	Permite operaciones de lectura y escritura	Estilo	Diseñado para tener un tiempo de respuesta rápido, una baja redundancia de datos y se normaliza



Universidad Autónoma Del Estado De México

Arquitectura de tres niveles

NIVEL INFERIOR: La base de datos de los servidores de Datawarehouse como el nivel inferior.

NIVEL MEDIO: El nivel intermedio de Datawarehouse es un servidor OLAP que se implementa utilizando el modelo ROLAP o MOLAP.

NIVEL SUPERIOR: El nivel superior es una capa de cliente front-end.



Universidad Autónoma Del Estado De México

Aplicaciones

Las aplicaciones OLTP tienen las características siguientes:

- Transacciones que implican pequeñas cantidades de datos
- Acceso indexado a los datos
- Numerosos usuarios
- Consultas y actualizaciones frecuentes
- Tiempos de respuesta rápidos



Universidad Autónoma Del Estado De México

Ejemplo

Sin embargo, la persona que completa el proceso de autenticación primero será capaz de obtener dinero.

En este caso, el sistema OLTP se asegura de que la cantidad retirada nunca será superior a la cantidad presente en el banco. La clave a tener en cuenta aquí es que los sistemas están optimizados para superioridad transaccional en lugar de análisis de datos



Universidad Autónoma Del Estado De México

MODELADO MULTIDIMENSIONAL

OLAP – BD Multidimensionales





Universidad Autónoma Del Estado De México

- Requiere de técnicas de diseño especializadas que se asemejan a los métodos tradicionales de diseño de bases de datos.
- En primer lugar se desarrolla una fase de diseño conceptual con el fin de obtener un **modelo multidimensional conceptual** para el almacén de datos independiente de la implementación.



Universidad Autónoma Del Estado De México

- Una de la herramientas basada en la plataforma de ECLIPSE y registradas por los autores de este texto desarrollada en la punta del grupo Lucentia.
- Basándose en estándares como UML, MDA. También puede destacar la herramienta WAND desarrollada en la Universidad de Bolonia.



Universidad Autónoma Del Estado De México

- El **lenguaje unificado de modelado (UML**, por sus siglas en inglés, *unified modeling language*) es el lenguaje de modelado de sistemas de software más conocido y utilizado en la actualidad; está respaldado por el object management group (omg).
- Model driven architecture® (MDA®) es un enfoque para el diseño, desarrollo e implementación de software liderado por el OMG. MDA proporciona directrices para estructurar las especificaciones de software que se expresan como modelos.



Universidad Autónoma Del Estado De México

MODELO MULTIDIMENSIONAL CONCEPTUAL

- Se debe acometer a nivel conceptual con el fin de reflejar de manera fidedigna y abstraer el diseño de la plataforma donde el almacén de datos será implementado.
- Debe representar aquellos elementos que permitan al usuario obtener información necesario para el apoyo a la toma de decisiones.
- Estos elementos o propiedades multidimensionales pueden ser estructurales o dinámicos.



Universidad Autónoma Del Estado De México

Propiedades multidimensionales estructurales

Consisten en la definición de estructuras de datos que faciliten el análisis multidimensional de los mismos.

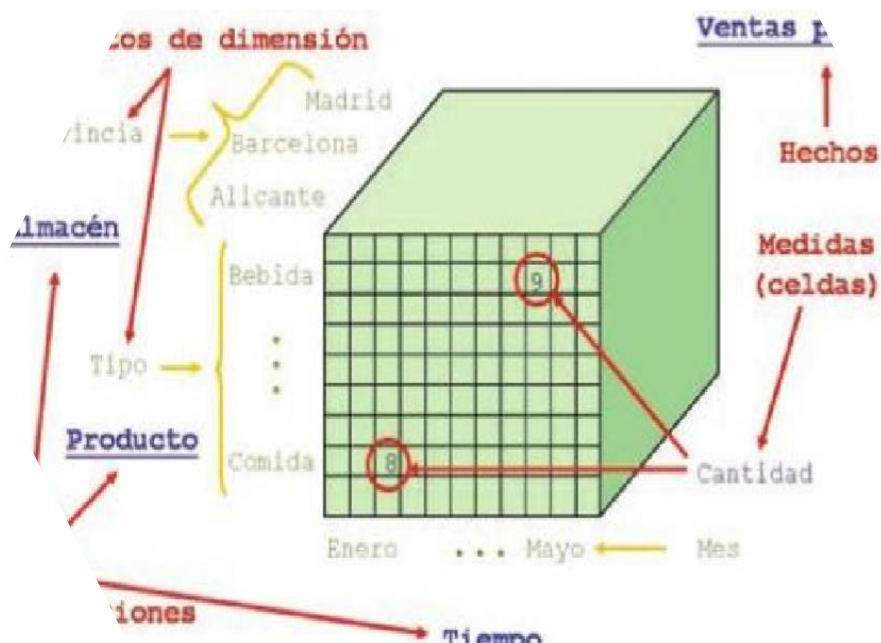
Un hecho contiene medidas interesantes que son el objeto de análisis, mientras que las dimensiones representan diferentes perspectivas para analizar dichas medidas.



Universidad Autónoma Del Estado De México

Representación de un modelo multidimensional

Se representa en forma de cubo o hipercubo o en su versión más sencilla, como tablas multidimensionales.



Ventas		Producto.Grupo = "Supermercado"				
		Comida		Bebida		
		Cong	Fresco	Refresco	Alcohol	
Almacén. comunidad = "Comunidad Valenciana"	Alicante	Albatera	100	200	300	400
		Elche	500	600	700	800
	Valencia	Burjasot	900	1000	1100	1200
		Cullera	1300	1400	1500	1600



Universidad Autónoma Del Estado De México

Propiedades multidimensionales dinámicas

Las propiedades dinámicas están relacionadas con la definición de los requisitos iniciales de usuario.

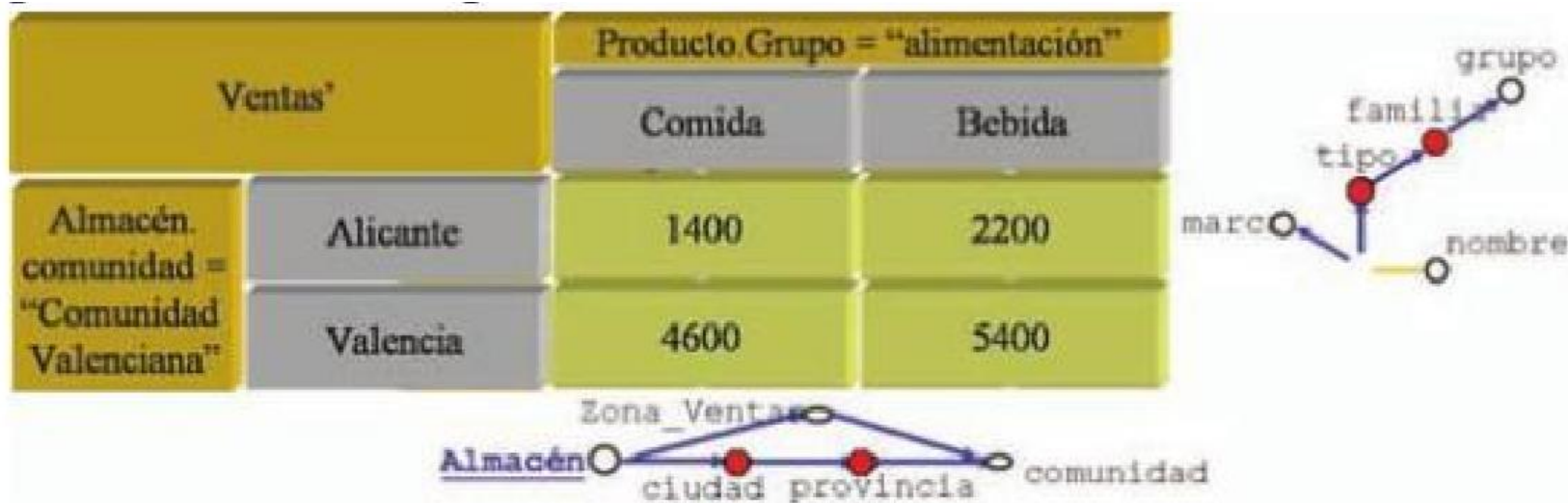
Para contestar algunos requisitos se deben utilizar un tipo de operaciones con el fin de interrogar a la parte estructural del modelo multidimensional.



Universidad Autónoma Del Estado De México

ROLL-UP

Esta operación permite agregar valores de medidas a lo largo de los niveles de jerarquías de clasificación de las dimensiones.





Universidad Autónoma Del Estado De México

DRILL-DOWN

Esta operación permite desagregar valores de medidas a lo largo de jerarquías de clasificación. Es la operación contraria al roll-up.

DRILL-ACROSS

Esta operación permite navegar de un hecho a otro mediante el uso de dimensiones comunes. Gracias a esta operación, se pueden consultar medidas de varios hechos en el mismo cubo.



Universidad Autónoma Del Estado De México

SLICE & DICE

Este par de operaciones permiten definir restricciones sobre niveles de jerarquías.

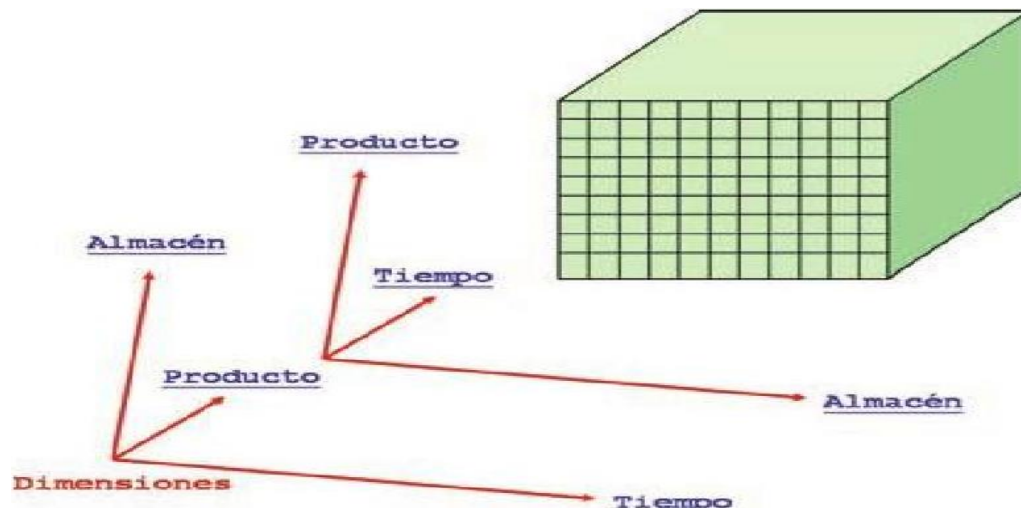
Ventas'			Producto.Grupo = "Alimentación"	
			Comida	
			Congelada	Fresca
Almacén. Comunidad = "Comunidad Valenciana"	Alicante	Albatera	100	200
		Elche	500	600



Universidad Autónoma Del Estado De México

PIVOTING

Esta operación permite reorientar la vista multidimensional de los datos, es decir, cambiar la distribución de filas o columnas. Algunos autores consideran también el intercambio de medidas y hechos como pivoting.





Universidad Autónoma Del Estado De México

DISEÑO LOGICO MULTIDIMENCIONAL

Toma como entrada un esquema conceptual y genera un esquema lógico relacional o multidimensional.

Esto tiene particular impacto en el caso de usarse bases relacionales, ya que las consultas de análisis de datos incluyen operaciones muy costosas para DBMS relacionales.



Universidad Autónoma Del Estado De México

Este modelo es independiente de cualquier herramienta y nos servirá para aclarar los conceptos generales.

Le permite a analistas y diseñadores más flexibilidad en el diseño, para lograr un mayor desempeño y optimizar la recuperación de la información, desde un punto de vista más cercano al usuario final.



Universidad Autónoma Del Estado De México

El modelo multidimensional produce una base de datos que es simple de navegar y encuestar.

Existen menor cantidad de tablas y relaciones en este, que en el modelo entidad-relación, el cual tiene cientos de tablas relacionadas entre sí y hay diferentes caminos para obtener una misma información, desde la perspectiva del usuario final resulta prácticamente inusable.



Universidad Autónoma Del Estado De México

Los tres componentes de un modelo de datos

Todo modelo de datos esta formado por tres componentes

- Estructura De Datos
- Operaciones Sobre Los Datos
- Restricciones De Integridad



Universidad Autónoma Del Estado De México

¿QUÉ ES MOLAP?

El OLAP multidimensional (MOLAP) es un OLAP clásico que facilita el análisis de datos mediante el uso de un cubo de datos multidimensional.

Los datos se calculan previamente, se resumen y se almacenan en un MOLAP (Una diferencia importante con respecto a ROLAP).



Universidad Autónoma Del Estado De México

¿Qué se puede hacer con un MOLAP?

Mediante un MOLAP, un usuario puede utilizar datos de vista multidimensional con diferentes facetas. El análisis de datos multidimensionales también es posible si se utiliza una base de datos relacional.

De este modo, sería necesario consultar datos de varias tablas.



Universidad Autónoma Del Estado De México

MOLAP tiene todas las combinaciones posibles de datos ya almacenados en un array multidimensional.

MOLAP puede acceder directamente a estos datos.



Universidad Autónoma Del Estado De México

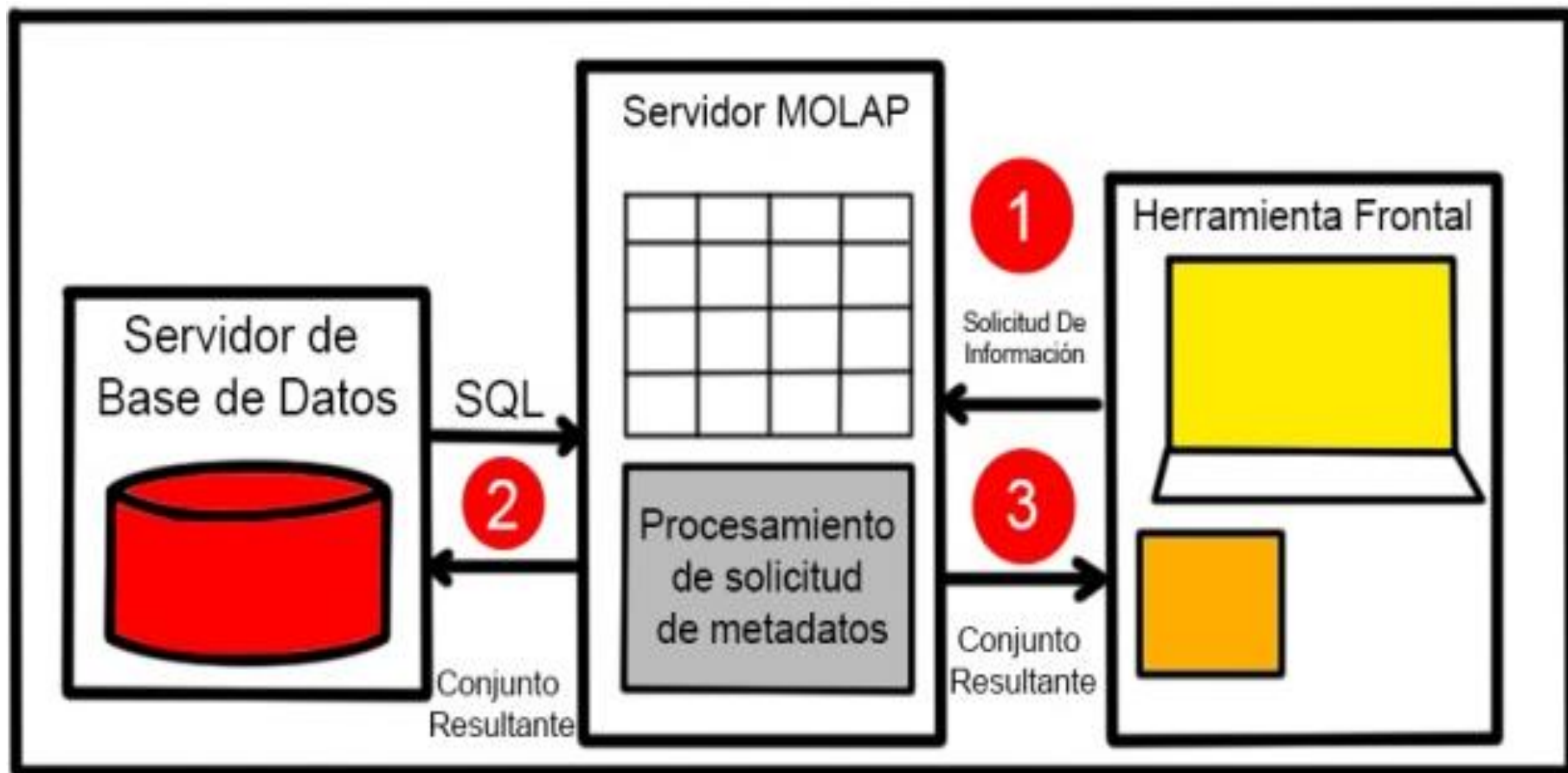
Arquitectura MOLAP

MOLAP Architecture incluye los siguientes componentes:

- Servidor de base de datos.
- Servidor MOLAP
- Herramienta frontal



Universidad Autónoma Del Estado De México





Universidad Autónoma Del Estado De México

1. Los informes de solicitud de usuario a través de la interfaz.
2. La capa lógica de aplicación de MDDB recupera los datos almacenados de la base de datos.
3. La capa lógica de la aplicación reenvía el resultado al cliente/usuario.



Universidad Autónoma Del Estado De México

Funcionamiento de la arquitectura

La arquitectura MOLAP lee principalmente los datos precompilados.

La arquitectura MOLAP tiene capacidades limitadas para crear agregaciones dinámicamente o para calcular resultados que no se han calculado ni almacenado previamente.



Universidad Autónoma Del Estado De México

Ejemplo

Un jefe contable puede ejecutar un informe que muestre la cuenta P/L corporativa o la cuenta P/L de una filial específica.

El MDDDB recuperaría las cifras pre-compiladas de Profit & Loss y mostraría ese resultado al usuario.



Universidad Autónoma Del Estado De México

Herramientas MOLAP

Nombre	Definición
ESSBASE	Herramienta de ORACLE que tienen una base de datos multidimensional.
YELLOWFIN	Herramientas de análisis empresarial para crear informes y paneles.
CLEAR ANALYTICS	Clear Analytics es una solución empresarial basada en Excel.
SAP BUSINESS INTELLIGENCES	Soluciones de análisis de negocios SAP



Universidad Autónoma Del Estado De México

Ventajas

- MOLAP puede gestionar, analizar y almacenar cantidades considerables de datos multidimensionales .
- Rendimiento de consultas rápido, gracias al almacenamiento optimizado, la indexación y el almacenamiento en caché.



Universidad Autónoma Del Estado De México

- MOLAP es más fácil para el usuario, por eso un modelo adecuado para usuarios sin experiencia.
- Todos los cálculos se generan previamente cuando se crea el cubo.



Universidad Autónoma Del Estado De México

Desventajas

- Las soluciones MOLAP pueden ser largas, especialmente en grandes volúmenes de datos.
- Los productos MOLAP pueden tener problemas al actualizar y consultar modelos cuando las dimensiones son más de 10.
- No es capaz de contener datos detallados.



Universidad Autónoma Del Estado De México

- Es menos escalable que ROLAP, ya que maneja sólo una cantidad limitada de datos.
- El MOLAP también introduce la redundancia de datos, ya que requiere un uso intensivo de recursos.



Universidad Autónoma Del Estado De México

CONCLUSIONES

- El procesamiento analítico en línea (OLAP) es una tecnología de análisis de datos que hace lo siguiente:
- Presenta una visión multidimensional lógica de los datos en el Data Warehouse. La visión es independiente de cómo se almacenan los datos.
- Comprende siempre la consulta interactiva y el análisis de los datos. Por lo regular la interacción es de varias pasadas, lo cual incluye la profundización en niveles cada vez más detallados o el ascenso a niveles superiores de resumen y adición.



Universidad Autónoma Del Estado De México

BIBLIOGRAFÍA

1. Krishna, R.,(2018).Aprender Almacenamiento de Datos en 1 Día.
2. Building the Data Warehouse, 4th Edition
3. Datawarehouse. Joaquina Martín Albo y Enrique Díaz
4. Diseño y explotación de Almacenes de datos. Juan C. Trujillo
5. Introducción a los Sistemas de Bases de Datos C. J. Date
6. DATA WAREHOUSING & DATA WAREHOUSE. Lic. Lisandra Fuentes Martínez. Ing. Leosvani Núñez García . The Data Warehouse Toolkit, 3rd Edition