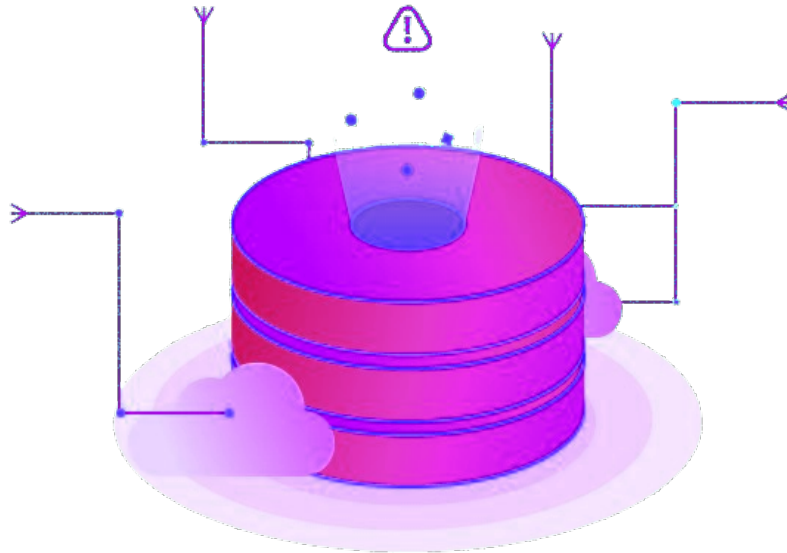


UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
FACULTAD DE CIENCIAS, 2024-I  
ALMACENES & MINERÍA DE DATOS



---

PRÁCTICA 04:  
*Creación de Cubos*

---

PROFESOR:  
Gerardo Avilés Rosas

AYUDANTE DE TEORÍA:  
Gerardo Uriel Soto Miranda

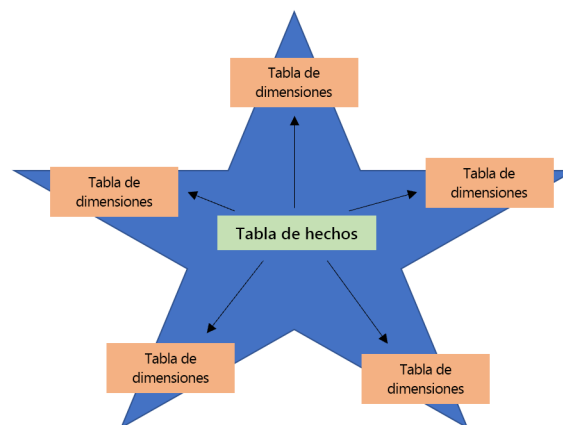
AYUDANTE DE LABORATORIO:  
Ricardo Badillo Macías

## Modelado de datos dimensionales:

Los modelos dimensionales son estructuras desnormalizadas diseñadas para recuperar datos de un almacén de datos. Están optimizados para realizar la operación "Seleccionar" se utilizan en el marco de diseño básico para construir almacenes de datos altamente optimizados y funcionales.

### Elementos de un modelo de datos dimensional

- **Hecho:** Los hechos son las medidas, métricas o hechos de tu proceso comercial.
- **Dimensión:** La dimensión proporciona el contexto que rodea un evento de proceso de negocio. En términos simples, proporciona el quién, qué, dónde de un hecho.
- **Tabla de hechos:** Una tabla de hechos es una tabla primaria en un modelo dimensional. La cual contiene: Mediciones/hechos y llave externa para tabla de dimensiones.
- **Tabla de Dimensiones:**
  - Una tabla de dimensiones contiene dimensiones de un hecho
  - Se unen a la tabla de hechos mediante una llave externa.
  - Las tablas de dimensiones son tablas desnormalizadas.
  - Los atributos de dimensión son las diversas columnas de una tabla de dimensiones.
  - Las dimensiones ofrecen características descriptivas de los hechos con la ayuda de sus atributos.
  - No se establece un límite establecido para el número de dimensiones.
  - La dimensión también puede contener una o más relaciones jerárquicas.



## Diseñar un modelo dimensional

La precisión en la creación de tu modelado dimensional determina el éxito de la implementación de tu almacén de datos. Estos son los pasos para crear un modelo dimensional eficaz capaz de describir el por qué, cuánto, cuándo, dónde, quién y qué de su proceso comercial.

- Identificar procesos de negocios
- Identificar la granularidad (nivel de detalle)
- Identificar dimensiones
- Identificar hechos.
- Construir Esquema.

## Identificación del proceso comercial

Esto es, identificar el proceso comercial real que debe abarcar un Almacenamiento de Datos. Esto podría ser Marketing, Ventas, Recursos Humanos, etc. según las necesidades de análisis de datos de la organización.

La selección del proceso de negocio también depende de la calidad de los datos disponibles para este proceso.

## Identifica la granularidad

La granularidad describe el nivel de detalle del problema. Es el proceso de identificar el nivel más bajo de información para cualquier tabla en tu almacén de datos. Si una tabla contiene datos de ventas de todos los días, debería ser una granularidad diaria.

Durante esta etapa, respondemos preguntas como:

- ¿Necesitamos almacenar todos los productos disponibles o solo algunos tipos de productos? Esta decisión se basa en los procesos comerciales seleccionados para el almacenamiento de datos.
- ¿Almacenamos la información de venta del producto en forma mensual, semanal, diaria o por hora? Esta decisión depende de la naturaleza de los informes solicitados por los ejecutivos.
- ¿Cómo afectan las dos opciones anteriores al tamaño de la base de datos?

## Identifica las dimensiones

Las dimensiones son sustantivos como fecha, tienda, inventario, etc. Estas dimensiones es donde se deben almacenar todos los datos. Por ejemplo, la dimensión de fecha puede contener datos como un año, mes y día de la semana. Un ejemplo de dimension sería:

- Dimensiones: producto, ubicación y hora.
- Atributos: Para producto: Clave de producto (Clave externa), Nombre, Tipo, Especificaciones.
- Jerarquías: Para ubicación: País, estado, ciudad, dirección, nombre.

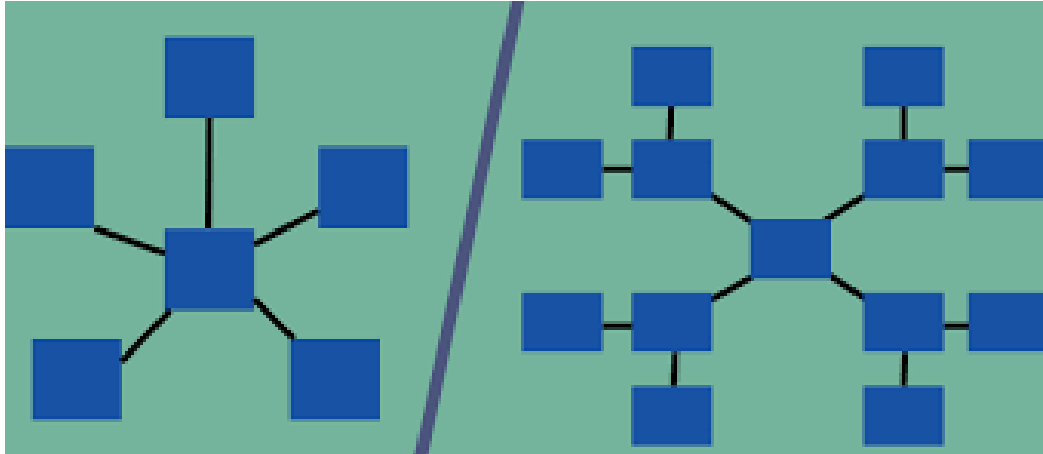
## Identifica el hecho

Este paso está asociado con los usuarios comerciales del sistema porque aquí es donde obtienen acceso a los datos almacenados en el almacén de datos. La mayoría de las filas de la tabla de hechos son valores numéricos como precio o costo por unidad, etc.

## Construir un esquema

En este paso, se implementará el modelo de dimensión. UN esquema no es más que la estructura de la base de datos (disposición de tablas) y hay dos esquemas populares.

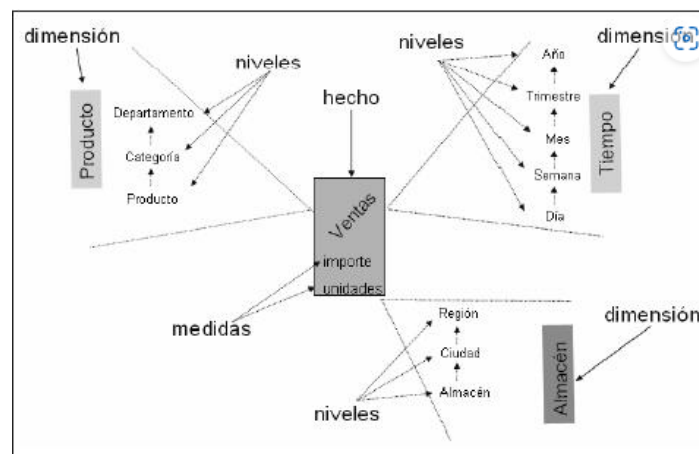
- **Esquema de estrella:** El diagrama se asemeja a una estrella, con puntos que irradian desde un centro. El centro de la estrella consiste en la tabla de hechos, y los puntos de la estrella son tablas de dimensiones.
- **Esquema de copo de nieve**  
El esquema de copo de nieve es una extensión del esquema de estrella. En un esquema de copo de nieve, cada dimensión se normaliza y se conecta a más tablas de dimensiones.



## Reglas para el modelado dimensional

Las siguientes son reglas básicas pero imprescindibles a la hora de crear tu modelado dimensional:

- Cargue datos atómicos en estructuras dimensionales.
- Construye modelos dimensionales en torno a los procesos empresariales.
- Asegúrate de que cada tabla de hechos tenga una tabla de dimensión de fecha asociada.
- Asegúrate de que todos los hechos en una sola tabla de hechos estén en el mismo nivel de detalle.
- Recuerda que es esencial almacenar etiquetas de informes y filtrar valores de dominio en tablas de dimensiones.
- Debes asegurarte de que las tablas de dimensiones utilicen una llave sustituta.
- Balancea continuamente los requisitos y las realidades para brindar una solución comercial que respalde su toma de decisiones.



## OLAP (Procesamiento analítico en línea)

La tecnología OLAP facilita el análisis de datos en línea en un almacén de datos, proporcionando respuestas a consultas analíticas complejas. OLAP es utilizado generalmente para ayuda en la toma de decisiones y presenta los datos a los usuarios a través de un modelo de datos intuitivo y natural. Con este estilo de presentación los usuarios finales pueden ver y entender con mayor facilidad la información de sus BD, lo que permite a las organizaciones reconocer el valor de sus datos.

Los cubos, las dimensiones y las jerarquías son la esencia de la navegación multidimensional del OLAP. Al describir y representar la información en esta forma, los usuarios pueden navegar intuitivamente en un conjunto complejo de datos.

Los cubos OLAP se pueden considerar como una ampliación de las 2 dimensiones de una hoja de cálculo, estos contienen información resumida de grandes bases de datos o Sistemas Transaccionales.

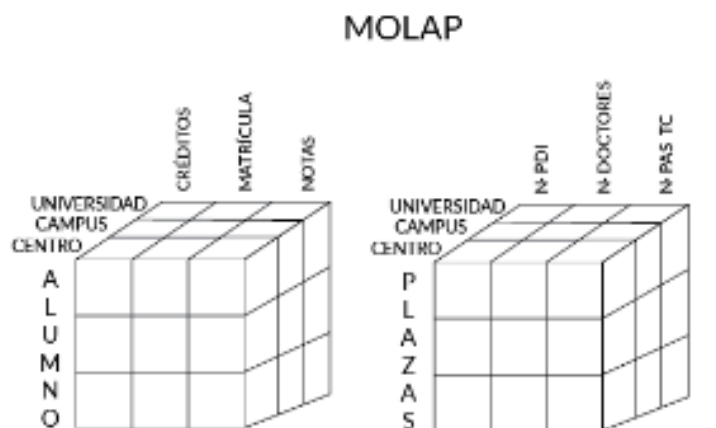
## Sistemas MOLAP

La arquitectura MOLAP, usa unas bases de datos multidimensionales para proporcionar el análisis, su principal premisa es que el OLAP está mejor implantado almacenando los datos multidimensionalmente.

El sistema MOLAP utiliza una arquitectura de dos niveles: la bases de datos multidimensionales y el motor analítico. La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato.

El nivel de aplicación es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través de la cual los usuarios finales visualizan los análisis OLAP.

La información procedente de los sistemas transaccionales se carga en el sistema MOLAP. Una vez cargados los datos en la BDM, se realiza una serie de cálculos para obtener datos agregados a través de las dimensiones del negocio, poblando la estructura de la BDMD.

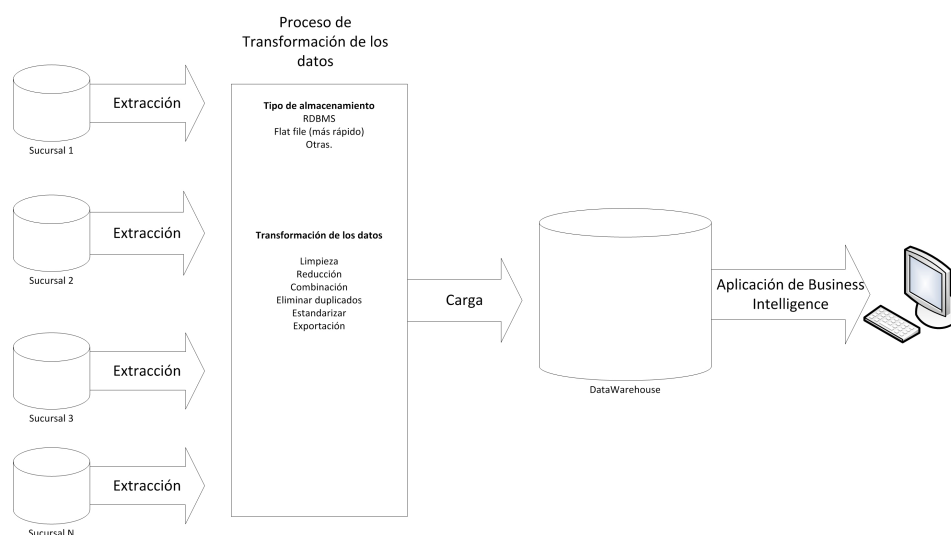


## SSAS (Analysis Services)

SSAS es una tecnología de Business Intelligence (BI), esta herramienta viene integrada con SQL Server.

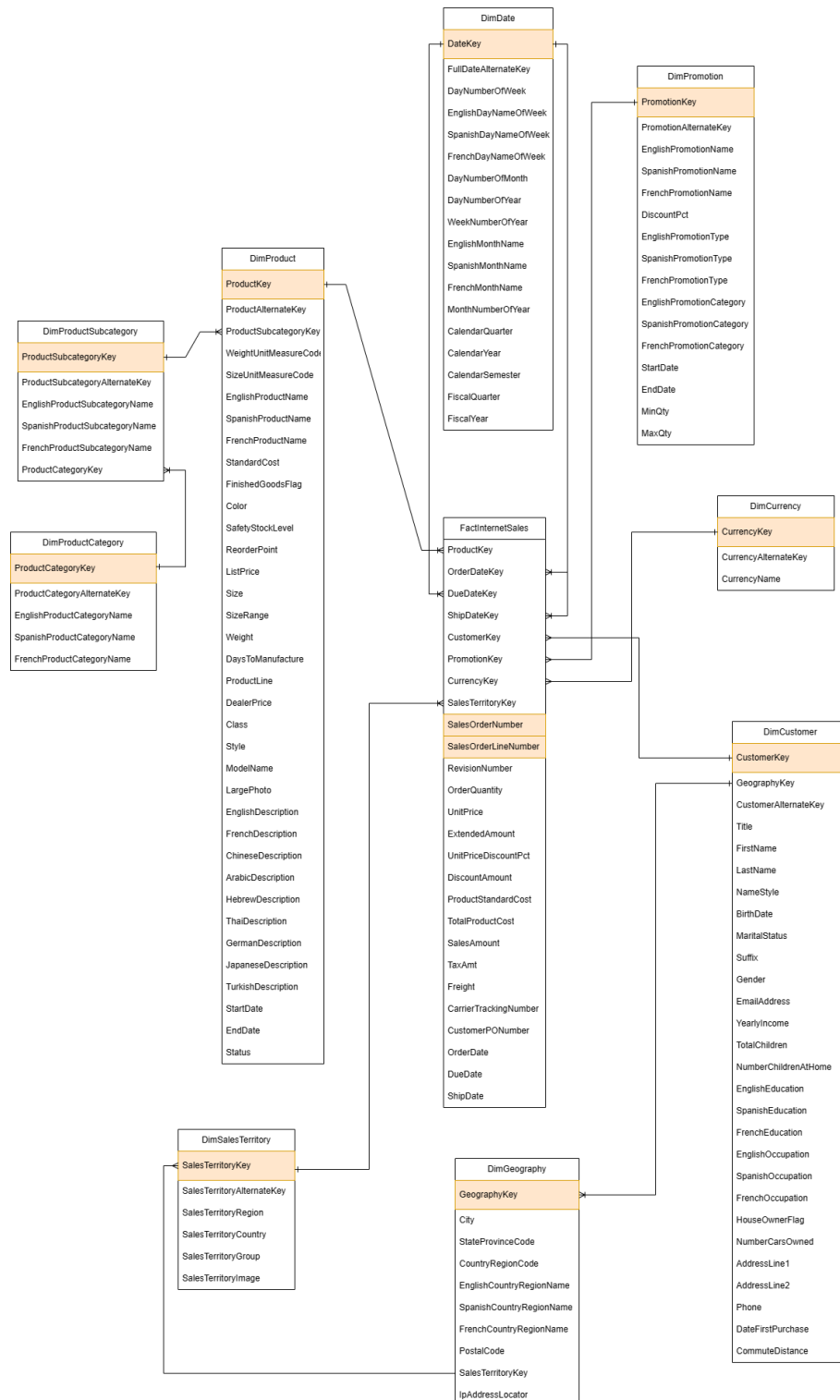
La herramienta funciona para crear informes analíticos rápidamente. Funciona con cubos diseñados para generar informes con un rendimiento óptimo utilizando información agregada almacenada en el cubo.

Un cubo en SSAS requiere un almacén de datos. Si se desea tener un cubo SSAS, el crear el almacén de datos es un proceso que requiere un poco de tiempo y recursos. Se puede utilizar SSIS para extraer la información de las fuentes de origen al almacén de datos o usar otras herramientas ETL (Extraer, Transformación y Carga).



## Actividades.

- i. Utilizaran **SSAS** para crear un cubo del siguiente diagrama multidimensional del almacén de datos “**AdventureWorksDW2012**” que se encuentra ya en el SQL Server de la maquina virtual.



- ii. Deberan **mostrar el proceso que hicieron para la creación del cubo**, y lo pondran en **Práctica04.pdf**, ademas deberan de especificar **el numero de dimensiones que se deberan de crear** y tambien especificar **cuantas jerarquías de conceptos hay**, justifiquen su respuesta. Deberan **seleccionar al menos 5 hechos y explicar** el porque los seleccionaron.
- iii. El proyecto debe de garantizar que se pueda implementar en la **Maquina Virtual**, y guardaran todo el proyecto en un archivo .zip, llamado **Cubo.zip**.
- iv. Una vez creado el cubo, deberan agregar a su **Práctica04.pdf una captura de consulta de datos**, en la cual se debe considerar **al menos un miembro de cada dimensión y un hecho**, esto se debera de hacer desde el explorador del cubo, en el proyecto.
- v. Contesta las siguientes preguntas:

Punto.	Pregunta.
1.	¿Que es un sistema de almacenamiento ROLAP?
2.	¿Cuáles son las principales diferencias que hay entre ROLAP y MOLAP?
3.	Explica a grandes rasgos, como se implementaria un sistema de almacenamiento ROLAP.
4.	¿En que se basa el sistema de almacenamiento HOLAP?



Figura 1: Actividades.



## Entregables.

Deberán subir un archivo con formato *zip* a *Google Classroom*, de acuerdo a lo indicado en los lineamientos de entrega. Debe de estar organizado de la siguiente manera, (suponiendo que el nombre del equipo que está entregando es *Dream Team* y los integrantes del equipo son los profesores del curso).

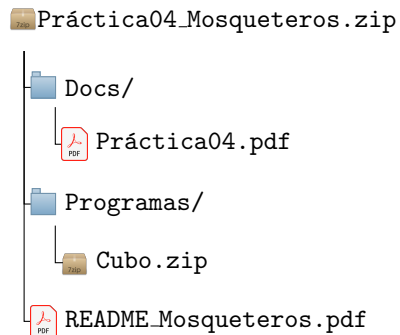


Figura 2: Entregables.

## Nota.

Para cualquier duda o comentario que pudiera surgirles al hacer este trabajo, recuerden que cuentan con la asignación de este entregable en el grupo de *Classroom*, en donde seguramente encontrarás las respuestas que necesites.



Figura 3: Nota.