**Data Warehouse.**

Un Data Warehouse (almacén de datos) es una base de datos centralizada que se utiliza para almacenar datos de múltiples fuentes en una estructura optimizada para el análisis y la toma de decisiones. El objetivo principal de un Data Warehouse es **permitir que los usuarios puedan consultar y analizar grandes cantidades de datos de manera eficiente y efectiva**, lo que puede incluir datos históricos y en tiempo real.

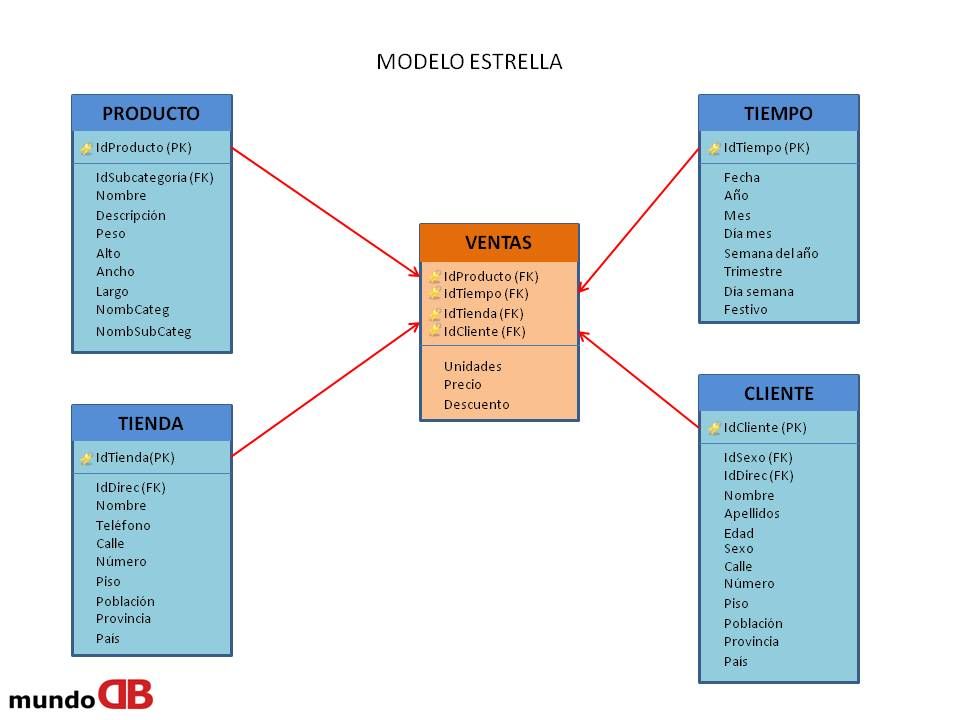
Además, los Data Warehouse suelen contar con herramientas y tecnologías que facilitan la integración de datos de diversas fuentes, su limpieza y transformación, así como su organización en esquemas específicos para el análisis. Esto permite que las empresas puedan tomar decisiones informadas basadas en datos precisos y confiables.

**Esquemas.**

Esquema de estrella

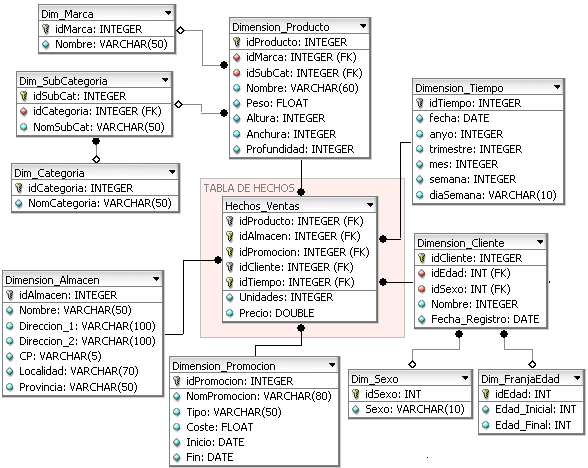
Se llama así porque la tabla central también llamada *tabla de hechos* está rodeada por varias tablas a las que se les denomina como *dimensiones*.

Las dimensiones también pueden formarse de otras tablas, es decir, (contener llaves foráneas de otras tablas).

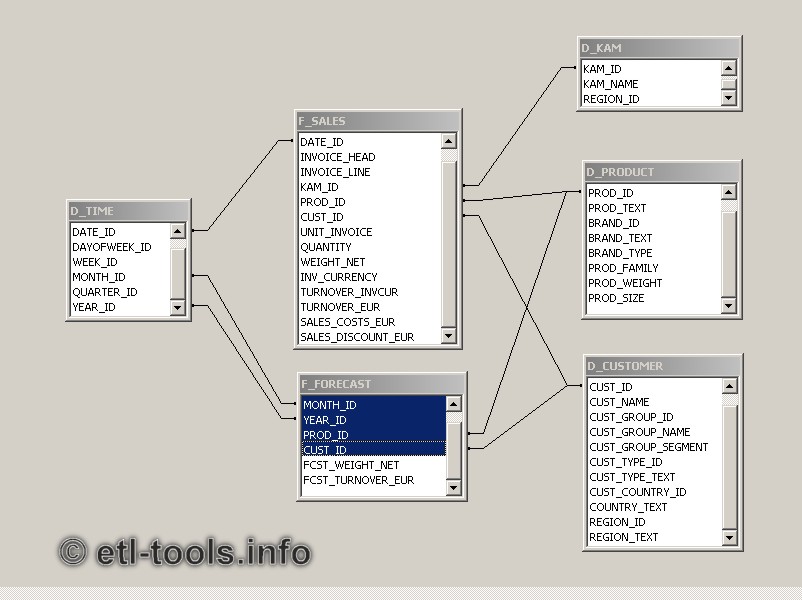
* La tabla de hechos recae sobre la segunda forma normal.
* *Contiene como atributos todas las llaves primarias de las entidades que la rodean*, (dimensiones).
* También puede contener sus propios atributos. De esta manera permite hacer consultas más rápidas y eficientes.

Esquema de copo de nieve

Es parecido al esquema de estrella, pero la diferencia radica en que unas dimensiones pueden tener conformarse de otras dimensiones para dividir la información. De esta manera ocupa menos espacio en la memoria, aunque es más complejo hacer consultas.

* Está diseñado para el mantenimiento de dimensiones.
* Cumple con la tercera forma normal.
* Ocupa menos espacio pero no tiene un rendimiento tan bueno, la razón es porque las consultas se vuelven más complejas de realizar.

Esquema de constelación

* Es más complejo que los dos anteriores, ya que contiene más de una tabla de hechos.
* Las dimensiones pueden estar repartidas entre las múltiples tablas de hechos.
* La ventaja de este sistema es su gran flexibilidad.
* Para generarla se sacrifica su facilidad y puede llegar a ser difícil darle mantenimiento en un futuro por el crecimiento de los datos.

Los esquemas anteriores logran la misma velocidad que el esquema estrella, (siempre y cuando se dividan los datos en dimensiones).

**Transacciones**

Una transacción es un proceso que no puede quedar a medias, (si empieza debe terminar), si por alguna razón se rechaza la transacción los cambios en la base de datos se deshacen. Todas las transacciones deben cumplir con ACID:

* **A**, atomicidad: Asegura que la operación se realice en su totalidad o que no se inicie.
* **C**, consistencia: Aseguro que solo se empieza aquello que se puede acabar.
* **I**, aislamiento: Asegura que una operación no puede afectar a otras.
* **D**, durabilidad: Asegura que una vez hecha la transacción, esta persistirá y no se podrá deshacer aunque falle el sistema.

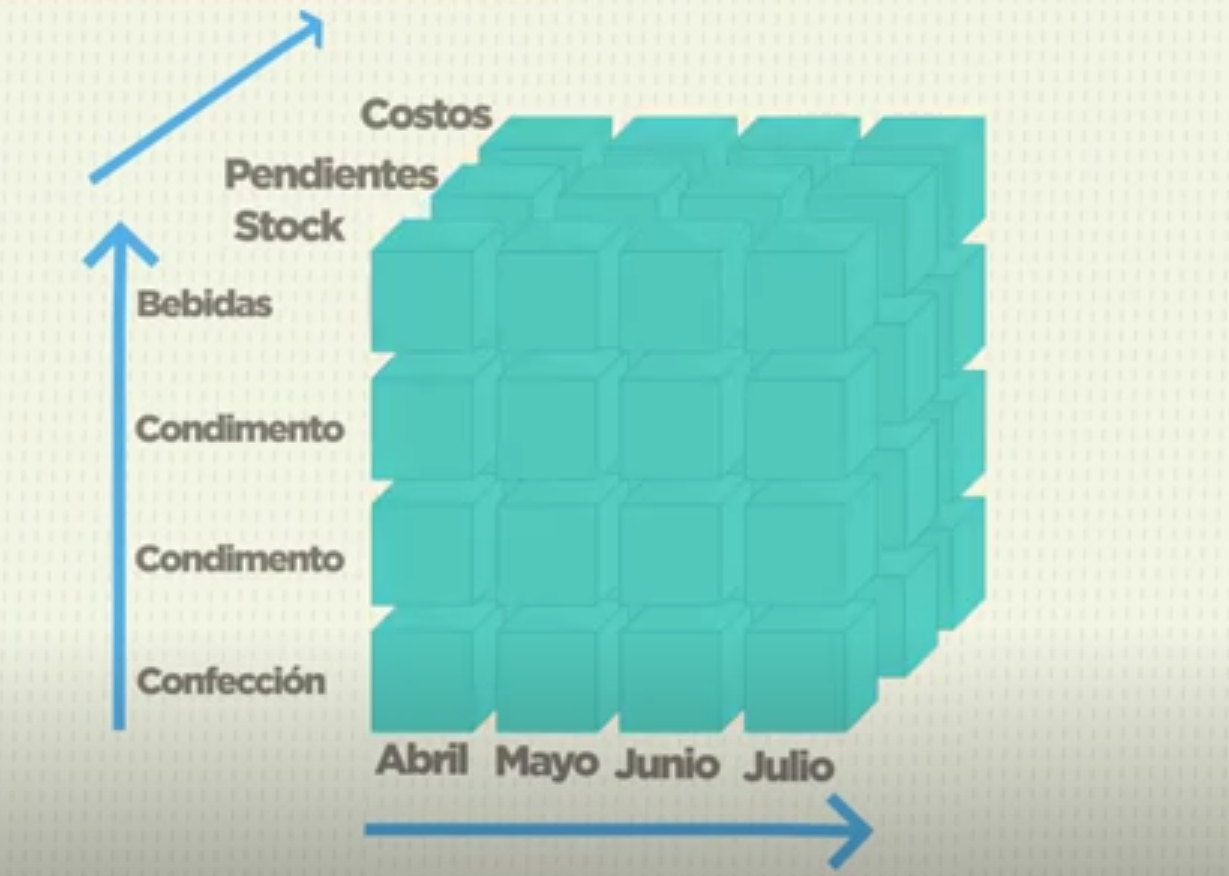
**ETL**

**Diseño del ETL**: El ETL (Extracción), (Transformación) y (Carga) de datos en una base de datos al Data Warehouse. Los Data Warehouse tienen una arquitectura distinta a una base de datos relacional, por ende los datos deben ser transformados para ser cargados.

El Data Warehouse debe estar diseñado de acuerdo a las bases de datos de dónde se extraerán los datos. Ya que si se extraen los datos de otra base que no fue contemplada, los datos no tendrán cabida.

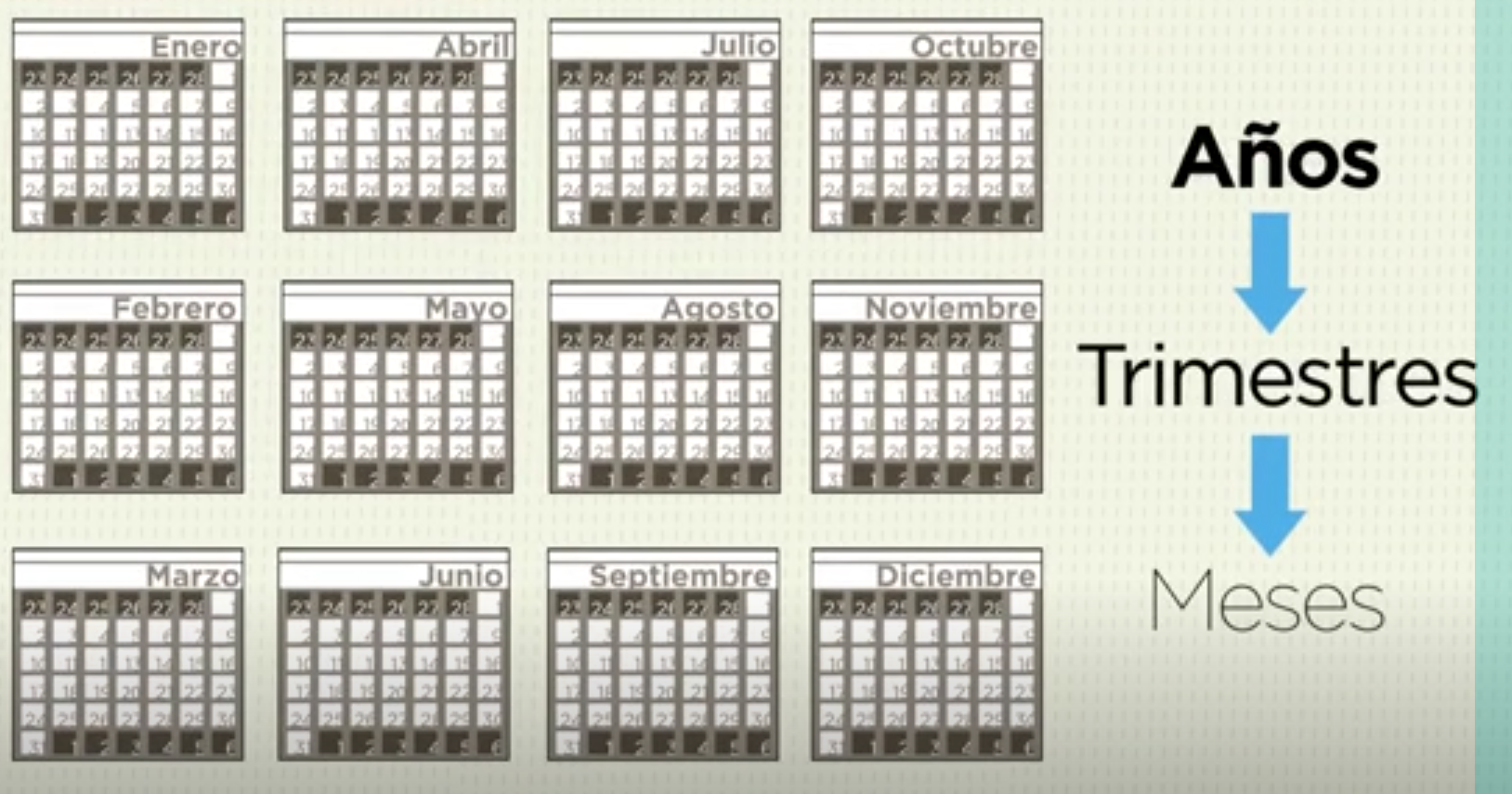
Una vez creado el diseño, se debe crear el Data Warehouse en el gestor con las configuraciones correspondientes. Para lograr un ETL en Data Warehouse, deben instalarse herramientas especiales de software como Bussines Intelligence.

**Cubos OLAP**

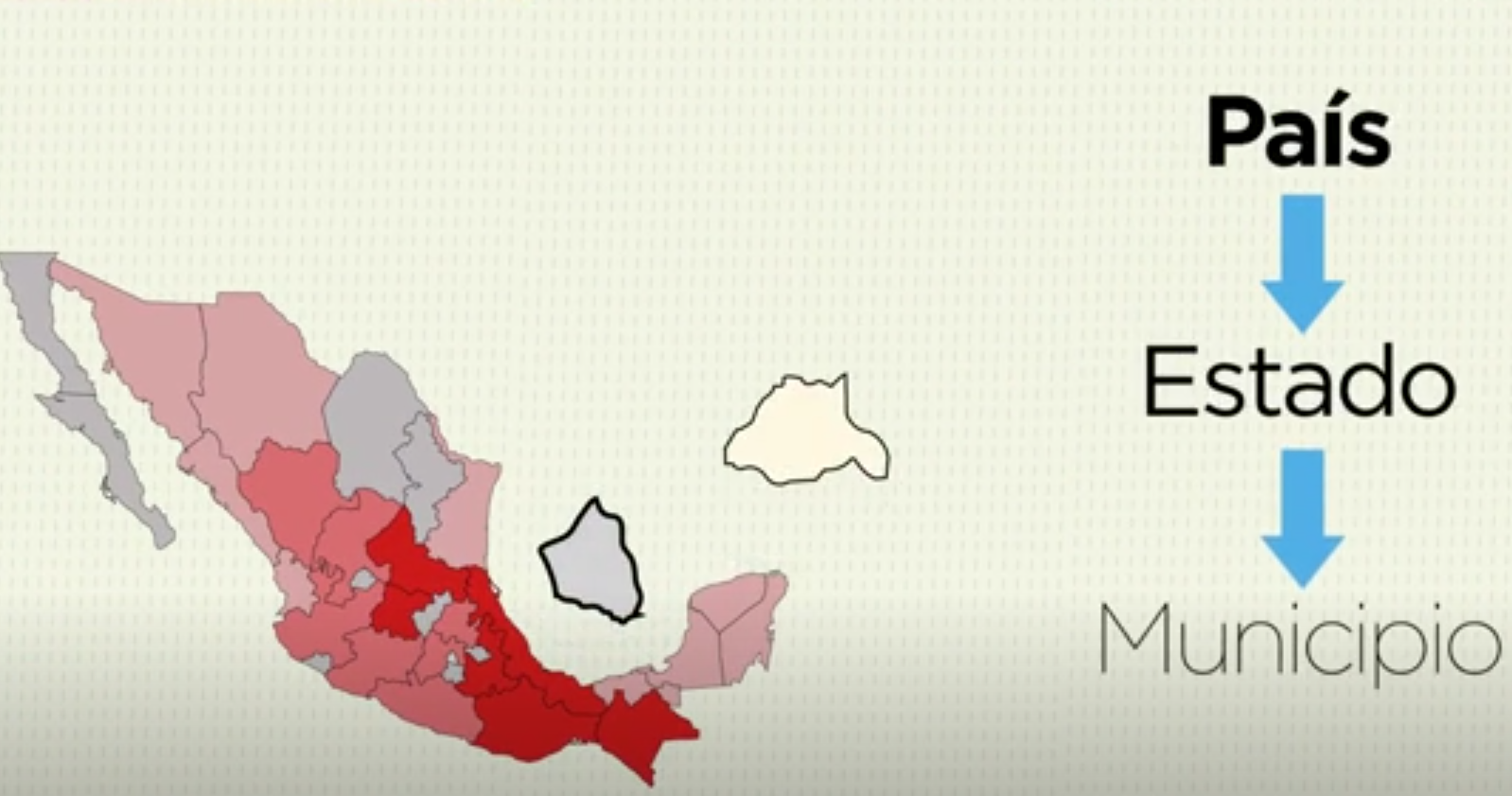
Las bases de datos relacionales son muy buenas para almacenar información, pero es difícil crear informes de datos utilizando tanta información. Los cubos OLAP, (Procesamiento Analítico Online) hacen más eficientes las bases de datos porque pre procesan la información y dan un fácil acceso a la misma.

Un cobo consta de tres dimensiones, por ende, si se busca información, esta se maneja de la siguiente manera: Como ejemplo: (Se busca la venta de productos de ciertas categorías en cierto tiempo).

Para obtener cierta información como por ejemplo: Las ventas totales de ciertos productos de la categoría Confección en el mes de Enero, se acceso a través de las intersecciones hasta obtener el resultado.

Otro ejemplo podría ser: El cubo se podría representar de la siguiente manera.

La dimensión Año es el cubo total, trimestres representan las filas del cubo y las celdas los meses.



Otro ejemplo podría ser la dimensión País, dónde tiene a estado como hijo y cada hijo tiene a la categoría Municipio.