**Esquemas Estrella y Copo de Nieve: Fundamentos y Contexto para Analistas de Datos**

Imagina que entras a una empresa y te entregan datos de clientes, productos y ventas en hojas distintas, con nombres duplicados y cifras que no cuadran. ¿Cómo construirías informes fiables? Antes de tocar un dashboard o hacer una consulta, necesitas ordenar y estructurar esos datos. Ese orden empieza por entender qué es una base de datos, cómo se modela y cómo se elige el esquema adecuado.

**Bases de datos: la materia prima**

Una base de datos es un sistema que guarda información de manera estructurada para poder almacenarla, actualizarla y consultarla. En las bases relacionales —las más habituales— los datos se organizan en tablas: cada fila es un registro (un cliente, un pedido) y cada columna es un atributo (nombre, fecha, cantidad). Para que todo funcione, cada tabla tiene reglas: una clave primaria (PK) para identificar unívocamente cada fila y una clave foránea (FK) para enlazarse con otras tablas. Las restricciones (NOT NULL, CHECK) ayudan a mantener la integridad.

En SQL existe además el término schema como “carpeta” que agrupa tablas (ventas.clientes). Pero en analítica la palabra “esquema” tiene otro sentido: se refiere al patrón de organización de las tablas, no al contenedor técnico.

**Modelado de datos:**

El modelado de datos es el proceso de analizar y diseñar cómo se van a organizar los datos de una organización. Es como el plano de un arquitecto antes de levantar un edificio. Con modelado respondemos: ¿Qué datos hay? (entidades). ¿Qué características tienen? (atributos). ¿Cómo se relacionan? (relaciones y reglas).

El modelado tiene tres niveles que se van refinando:

* Conceptual: representación general del negocio, sin tablas ni SQL.
* Lógico: traducción a tablas y columnas, con PK y FK, pero sin tipos específicos.
* Físico: implementación real en SQL con CREATE TABLE, tipos concretos e índices.

Este modelado no es exclusivo de bases relacionales. También se modelan datos en NoSQL, grafos o data lakes. Lo importante es pensar y diseñar antes de construir. Y, dependiendo de para qué vas a usar los datos, el modelo tomará distintas formas.

**Normalización y desnormalización:** orden vs rapidez

En sistemas operativos que registran ventas, pagos o reservas, nos importa la integridad y las escrituras rápidas. Por eso usamos normalización: dividimos los datos en varias tablas relacionadas para evitar duplicidad y anomalías.

En cambio, en sistemas analíticos nos interesa leer y agregar rápido. Ahí usamos desnormalización: duplicamos atributos descriptivos en una sola tabla o en pocas tablas para acelerar consultas y simplificar dashboards. Esto es la base de los modelos dimensionales.

**OLTP y OLAP**

Cuando un negocio ya tiene datos, aparecen dos necesidades diferentes:

* OLTP (Online Transaction Processing): sistemas que registran operaciones del día a día. Bases normalizadas, diseñadas para integridad y velocidad en escrituras. Ej.: un banco registrando transferencias, un ecommerce registrando pedidos.
* OLAP (Online Analytical Processing): sistemas que analizan grandes volúmenes de datos históricos. Bases desnormalizadas/dimensionales, diseñadas para lecturas y agregaciones rápidas. Ej.: un dashboard que muestra ventas por región y tendencias.

Como analistas de datos, normalmente no se conectan al OLTP, sino a la capa analítica (data warehouse, data mart) pensada para consultas y dashboards.

**Hechos y dimensiones:**

En analítica de datos, todo gira alrededor de hechos y dimensiones:

* **Hecho:** un evento medible. Ej.: una venta, un clic, una transferencia.
* **Dimensión:** el contexto de ese hecho. Ej.: cliente, producto, fecha, región.

**Ejemplo:**

**Hecho:** Se vendieron 3 camisetas por $50 el 25/09/2025.

**Dimensiones**:

* Cliente = Juan Pérez
* Producto = Camiseta
* Fecha = 25/09/2025
* Región = Atlántico

Cuando entiendes esta división, estás preparado para ver cómo se organizan en los esquemas estrella y copo de nieve.

**Esquema en analítica como diseño lógico**

En el contexto del modelado de datos y la analítica, “esquema” no significa “carpeta de SQL”, sino el patrón de organización de las tablas en una base orientada a análisis. Es el plano lógico que dice: qué tablas existen, qué columnas tienen, cómo se unen y qué clave conecta con qué.

Es un mapa de la estructura de los datos que sirve para diseñar data warehouses o data marts eficientes. Y los dos patrones más conocidos son el esquema estrella y el esquema copo de nieve.

**Esquema Estrella y Esquema Copo de Nieve: patrones para análisis**

Una vez que tienes claro qué hechos y dimensiones hay y que estás en un contexto OLAP, puedes presentar los patrones de organización:

**Esquema Estrella**

La tabla de hechos en el centro y alrededor todas las dimensiones en tablas independientes. Se llama “estrella” porque gráficamente parece una estrella: un núcleo y puntas. Las dimensiones suelen estar desnormalizadas (incluyen directamente atributos como nombre de categoría, región…), lo que hace las consultas más simples y rápidas.

**Esquema Copo de Nieve**

Similar al estrella, pero con dimensiones normalizadas. Por ejemplo, en vez de tener la categoría dentro de DimProducto, creas una tabla DimCategoría aparte y la unes. Esto ahorra espacio y evita duplicidad en dimensiones muy grandes, pero añade más JOINs en las consultas. Gráficamente se ve como un copo con ramificaciones.

**¿Cuándo usar cada uno?**

* Para dashboards y análisis rápidos (Power BI, Tableau, Looker): Esquema Estrella es la opción estándar.
* Para data warehouses muy grandes con dimensiones cambiantes o muy detalladas: Copo de Nieve puede ser mejor para mantenimiento y espacio.
* En sistemas transaccionales (OLTP) no se usan estos esquemas; ahí va el modelo normalizado.