# Investigación – Modelo de Datos

## ¿Qué es un Modelo de Datos?

Los modelos de datos definen cómo se modela la estructura lógica de una base de datos. Los modelos de datos son entidades fundamentales para introducir la abstracción en una base de datos.

Los modelos de datos definen cómo los datos se conectan entre sí y cómo se procesan y almacenan dentro del sistema.

El primer modelo de datos fue el modelo de datos planos, donde todos los datos utilizados se mantendrían en el mismo plano.

Típicamente un modelo de datos permite describir:

* **Las estructuras de datos de la base:** El tipo de los datos que hay en la base y la forma en que se relacionan.
* **Las restricciones de integridad:** Un conjunto de condiciones que deben cumplir los datos para reflejar la realidad deseada.
* **Operaciones de manipulación de los datos:** típicamente, operaciones de agregado, borrado, modificación y recuperación de los datos de la base.

Otro enfoque es pensar que un modelo de datos permite describir los elementos de la realidad que intervienen en un problema dado y la forma en que se relacionan esos elementos entre sí.

No hay que perder de vista que una base de datos siempre está orientada a resolver un problema determinado, por lo que los dos enfoques propuestos son necesarios en cualquier desarrollo de software.

Un modelo de datos puede ser concreto o abstracto, y están representados por la notación de **modelado de datos**, que a menudo se presenta en formato gráfico.

También ayudan a evitar la redundancia de datos. La información almacenada en los modelos de datos es de gran importancia para las empresas porque dicta las relaciones entre las tablas de la base de datos, las claves externas y los eventos involucrados.

## Modelado de Datos

El modelado de datos es una manera de estructurar y organizar los datos para que se puedan utilizar fácilmente por las bases de datos. Los datos no estructurados se pueden encontrar en los documentos de procesamiento de texto, mensajes de correo electrónico, archivos de audio o vídeo, y programas de diseño.

El modelado de datos no quiere estos datos "crudos" sino que el modelado de datos quiere que todos los datos se presenten en un paquete bonito, limpio para el procesamiento de una base de datos. Así que, en cierto modo, el modelado de datos se refiere a cómo se ven los datos.

El modelado de datos se utiliza habitualmente en combinación con un sistema de gestión de base de datos. Los datos que se han modelado y preparado para este sistema se pueden identificar de varias maneras, como de acuerdo con lo que representan, o cómo se relacionan con otros datos. La idea es hacer de los datos tan presentables como sea posible, para que el análisis y la integración se pueda hacer con tan poco esfuerzo como sea necesario.

También podemos pensar en el modelado de datos como las instrucciones para la construcción de una base de datos. Concéntrese en la palabra modelo y entenderá a qué nos referimos. Para hacer una buena base de datos, tendrá que seguir un modelo como un medio para su fin deseado.

## Tipos de Modelos de Datos

Una opción bastante usada a la hora de clasificar los modelos de datos es hacerlo de acuerdo con el nivel de abstracción que presentan:

* **Modelos de Datos Conceptuales**

Son los orientados a la descripción de estructuras de datos y restricciones de integridad. Se usan fundamentalmente durante la etapa de Análisis de un problema dado y están orientados a representar los elementos que intervienen en ese problema y sus relaciones. El ejemplo más típico es el Modelo Entidad-Relación.

Las características del modelo conceptual de datos incluyen:

* + Entidades importantes y las relaciones entre ellas.
  + No se especifica ningún atributo.
  + No se especifica ninguna clave principal.
* **Modelos de Datos Lógicos**

Son orientados a las operaciones más que a la descripción de una realidad. Usualmente están implementados en algún Manejador de Base de Datos. El ejemplo más típico es el Modelo Relacional, que cuenta con la particularidad de contar también con buenas características conceptuales (Normalización de bases de datos).

Las características de un modelo de datos lógicos incluyen:

* + Entidades y relaciones entre ellos.
  + Todos los atributos para cada entidad están especificados.
  + La clave principal para cada entidad está especificada.
  + Se especifican las claves externas (claves que identifican la relación entre diferentes entidades).

* **Modelos de Datos Físicos**

Son estructuras de datos a bajo nivel implementadas dentro del propio manejador. Ejemplos típicos de estas estructuras son los Árboles B+, las estructuras de Hash, etc.

Un modelo de base de datos física muestra todas las estructuras de tabla, incluidos el nombre de columna, el tipo de datos de columna, las restricciones de columna, la clave principal, la clave externa y las relaciones entre las tablas.

Las características de un modelo de datos físicos incluyen:

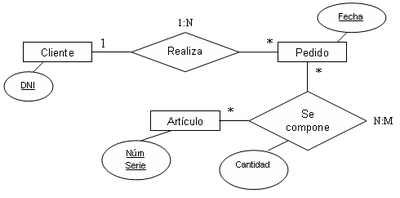
* + Especificación de todas las tablas y columnas.
  + Las claves externas se usan para identificar relaciones entre tablas.
  + La desnormalización puede ocurrir según los requisitos del usuario.

## Desarrollo de un Modelo de Datos

Para un desarrollador de aplicaciones, es fundamental conocer los fundamentos del modelado de datos para que no solo puedan leer modelos de datos sino también trabajar eficazmente con los administradores de agile que son responsables de los aspectos orientados a los datos de su proyecto.

### Identificar tipos de entidades

Un tipo de entidad, también llamado simplemente entidad (no es exactamente una terminología precisa, pero muy común en la práctica), es similar conceptualmente al concepto de una clase de orientación a objetos: un tipo de entidad representa una colección de objetos similares.

Un tipo de entidad podría representar una colección de personas, lugares, cosas, eventos o conceptos. Los ejemplos de entidades en un sistema de entrada de pedidos incluirían Cliente, Pedido, y Artículo.

La diferencia entre una clase y un tipo de entidades que las clases tienen datos y comportamiento, mientras que los tipos de entidad solo tienen datos.

Idealmente, una entidad debería ser normal, y de cohesión en el modelado de datos. Una entidad normal representa un concepto, al igual que una clase cohesiva modela un concepto.

Por ejemplo, cliente y pedido son claramente dos conceptos diferentes, por lo tanto, tiene sentido modelarlos como entidades separadas.

### Identificar atributos

Cada tipo de entidad tendrá uno o más atributos de datos. Por ejemplo, en la Figura anterior se ve que Cliente tiene atributos como Nombre y Teléfono. Y en la tabla de entidades, las entidades tienen columnas de datos correspondientes a sus atributos (una columna es la implementación de un atributo de datos dentro de una base de datos relacional).

Los atributos también deben ser coherentes desde el punto de vista de su dominio, algo que a menudo es una decisión. En el diagrama decidimos que queríamos modelar el hecho de que las personas tenían tanto Nombre y Teléfono (también pudo Nombre, Apellido y Teléfono).

Obtener el nivel de detalle correcto puede tener un impacto significativo en sus esfuerzos de desarrollo y mantenimiento. Refactorizar una sola columna de datos en varias columnas puede ser difícil, la refacturación de bases de datos se describe en detalle en refactorización de la base de datos, aunque especificar demasiado un atributo (por ejemplo, tener tres atributos para el código postal cuando solo se necesita uno) puede generar una sobrecarga del sistema y, por lo tanto, incurrirá en mayores costos de desarrollo y mantenimiento de los que realmente necesita.

### Aplicar convenciones de nomenclatura

Su organización debe tener estándares y pautas aplicables al modelado de datos, algo que debería poder obtener de los administradores de su empresa (si no existen, debe presionar para que se existan algunos).

Estas directrices deben incluir convenciones de nomenclatura para el modelado lógico y físico, las convenciones de nomenclatura lógica deben centrarse en la legibilidad humana, mientras que las convenciones de nomenclatura física reflejarán consideraciones técnicas.

Puede ver claramente que se aplicaron diferentes convenciones de nomenclatura en la figura Cliente-Pedido (1, \*).

La idea básica es que los desarrolladores deben aceptar y seguir un conjunto común de estándares de modelado en un proyecto de software. Así como hay un valor en seguir las convenciones de codificación comunes, el código de limpieza que sigue las pautas de codificación elegidas es más fácil de comprender y evolucionar que el código que no lo hace, y existe un valor similar en las siguientes convenciones de modelado comunes.

### Identificar relaciones

En el mundo real, las entidades tienen relaciones con otras entidades. Por ejemplo, los clientes realizan pedidos, los clientes viven en direcciones, y las líneas de pedido forman parte de los pedidos. Realizar, vivir y formar parte de todos los términos que definen las relaciones entre las entidades.

Las relaciones entre entidades son conceptualmente idénticas a las relaciones (asociaciones) entre objetos.

### Aplicar patrones de Modelos de Datos

Algunos modeladores de datos aplicarán patrones de modelos de datos comunes, el modelo de datos de David Hay es la mejor referencia sobre el tema, así como los desarrolladores orientados a objetos aplicarán patrones de análisis (Fowler 1997, Ambler 1997) y patrones de diseño (Gamma et al., 1995).

Los patrones del modelo de datos son conceptualmente los más cercanos a los patrones de análisis porque describen soluciones para problemas de dominio común. Los modelos de Hay son una referencia muy buena para cualquier persona involucrada en el modelado de nivel de análisis, incluso cuando se toma un enfoque de objetos en lugar de un enfoque de datos porque sus patrones modelan estructuras comerciales de una amplia variedad de dominios comerciales.

### Asignar llaves

Hay dos estrategias fundamentales para asignar claves a las tablas. En primer lugar, podría asignar una clave natural que sea uno o más atributos de datos existentes que sean exclusivos del concepto de negocio.

La tabla Cliente puede tener unas claves candidatas, en este caso DNI o número de identificación nacional. En segundo lugar, podría introducir una nueva columna, llamada clave sustituta, que es una clave que no tiene ningún significado comercial.

### Normalizar para reducir la redundancia

La normalización de datos es un proceso en el que los atributos de los datos dentro de un modelo de datos se organizan para aumentar la cohesión de los tipos de entidades.

En otras palabras, el objetivo de la normalización de datos es reducir e incluso eliminar la redundancia de datos, una consideración importante para los desarrolladores de aplicaciones porque es increíblemente difícil almacenar objetos en una base de datos relacional que mantiene la misma información en varios lugares.

## Ejemplo

