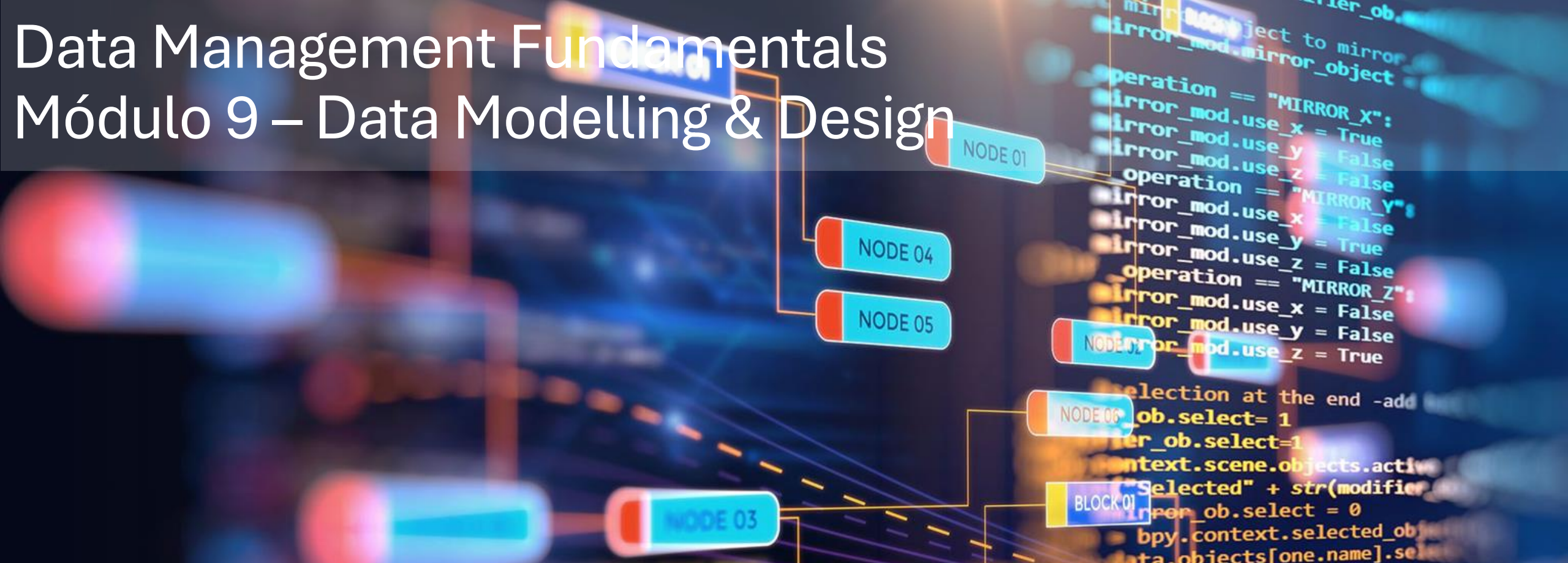


Data Management Fundamentals

Módulo 9 – Data Modelling & Design



GOLD SPONSORS



SILVER SPONSORS

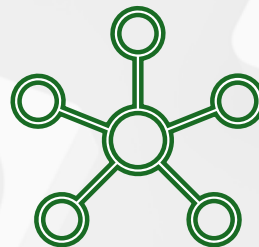


Entidades académicas



Modelado de datos

- Es el proceso de descubrir, analizar y definir el alcance de los requerimientos de datos, con el objetivo de representarlos y comunicarlos de forma precisa en una forma llamada modelo de datos
- Los modelos de datos describen y permiten a una organización entender sus activos de datos

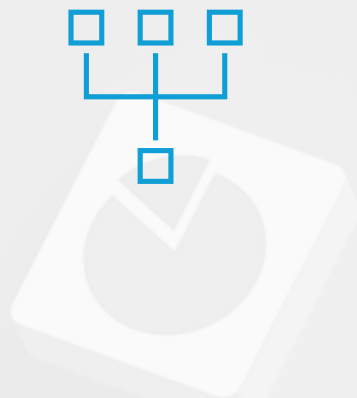


Ventajas de los modelos de datos

- Proporcionan un vocabulario común en torno a los datos
- Capturan y documentan el conocimiento explícito sobre los datos y sistemas de una organización
- Sirven como herramienta principal de comunicación durante los proyectos
- Proporcionan el punto de partida para la personalización, la integración o incluso la sustitución de una aplicación

Objetivo

Confirmar y documentar la comprensión de las diferentes perspectivas, **alineando** las aplicaciones con los **requisitos empresariales actuales y futuros**, y creando la base del éxito para iniciativas de amplio alcance como la gestión de datos maestros y los programas de gobierno de datos.



Ventajas

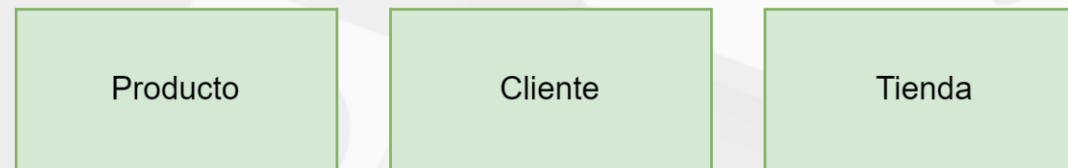
- **Formalización:** Poseer una definición concisa de las estructuras y relaciones de datos permite evaluar cómo los datos se ven afectados por la implementación de reglas de negocio. Esta documentación, además, facilita el consumo de datos
- **Definición del alcance:** Ayuda a explicar los límites del contexto de los datos y la implementación de aplicaciones
- **Retención de conocimientos/documentación:** Preserva la memoria de la empresa al capturar los conocimientos de forma explícita. Se trata de un mapa reutilizable para ayudar a entender la estructura de datos en el entorno

Tipos de datos modelados

- **Información sobre la categoría:** Datos utilizados para clasificar y asignar tipos a las cosas como clientes clasificados por categorías de mercado o productos clasificados por modelo
- **Información de recursos:** Perfiles básicos de los recursos necesarios para realizar procesos operativos tales como producto o cliente
- **Información de eventos de negocios:** Los datos creados durante los procesos operativos. Por ejemplo, pedidos de clientes o facturas de proveedor
- **Información de detalle de la transacción:** Información detallada de transacciones producida en sistemas de punto de venta o redes sociales. Son datos muy variables y abundantes que habitualmente se tratan desde una perspectiva de Big Data

Entidades

- Son el metadato básico, son un algo sobre lo que la organización recaba datos
- Puede considerarse como la respuesta a una pregunta fundamental: Quién, Qué, Cuándo, Dónde, Por Qué, Cómo o a una combinación de las mismas
- Suele representarse como un rectángulo



Categorías de entidades típicas

Categoría	Definición	Ejemplos
Quién	Persona u organización de interés. ¿Es decir, <i>quién</i> es importante para el negocio? Un ' <i>quién</i> ', está a menudo asociado con una generalización de partido, o rol como cliente o proveedor. Personas u organizaciones pueden tener múltiples roles o ser incluidas en múltiples partes.	Empleado, Paciente, Jugador, Sospechoso, Cliente, Vendedor, Alumno, Pasajero, Competidor, Autor
Qué	Producto o servicio de interés para la empresa. A menudo se refiere a lo que la organización hace o qué servicio ofrece. Es decir, ¿ <i>qué</i> es importante para el negocio? Atributos por categorías, tipos, etc. son muy importantes aquí.	Producto, Servicio, Materia Prima, Bien Acabado, Curso, Fotografía, Canción, Libro
Cuándo	Intervalo de tiempo o calendario de interés para la empresa. Es decir, ¿ <i>cuándo</i> está el negocio en funcionamiento?	Tiempo, Fecha, Mes, Trimestre, Año, Calendario, Semestre, Período Fiscal, Minuto, Hora de Salida
Dónde	Ubicación de interés para la empresa. Ubicación puede hacer referencia a lugares como lugares electrónicos. Esto es ¿ <i>dónde</i> se conduce el negocio?	Dirección Postal, Punto de Distribución, URL del sitio web, Dirección IP
Por qué	Evento o transacción de interés para la empresa. Estos eventos mantienen el negocio a flote. Esto es ¿ <i>Por qué</i> está el negocio en el negocio?	Orden, Retorno, Denuncia, Retiro, Depósito, Complemento, Consulta, Tratado, Demanda
Cómo	Documentación de los eventos de interés para la empresa. Los Documentos proporcionan la evidencia de que los eventos ocurrieron, como una Orden de compra, registrar un evento de Orden. Es decir, ¿ <i>Cómo</i> sabemos que un evento ocurrió?	Factura, Contrato, Acuerdo, Cuenta, Orden de Compra, Multa por exceso de velocidad, Remisión, Confirmación de comercio
Medición	Cuenta, sumas, etc. de las otras categorías (qué, dónde) sobre los puntos en el tiempo (Cuándo).	Ventas, Cuenta de artículos, Pagos, Saldo

Definición de entidades

Al tratarse de metadatos fundamentales para el valor aportado al negocio por el modelo de datos, deben estar definidos de forma clara y con gran calidad:

- **Claridad:** La definición debe ser fácil de leer y comprender. Se recomienda utilizar frases simples y bien escritas sin acrónimos o términos ambiguos inexplicados. Evitar el uso de expresiones como “a veces” o “normalmente”
- **Precisión:** La definición debe ser una descripción precisa y correcta de la entidad. Deben ser revisadas por expertos de negocio para asegurar que sean precisas
- **Compleitud:** Todas las partes de la definición están detalladas. Por ejemplo, al definir código, se incluyen ejemplos de los valores del código. Al definir un identificador, se incluye en la definición el alcance de la unicidad

Atributos

Son una **propiedad que identifica, describe o mide a una entidad.**

Se suele representar como una lista bajo la entidad.

Producto
idproducto
codigobarras
categoria
marca
fechaalta
fechabaja
idproveedor
costecompra
precioventa

Dominios

Es el **conjunto completo de valores posibles** de un atributo:

- Proporciona medios para estandarizar las características de los atributos
- Un atributo sólo puede contener valores válidos definidos en su dominio
- Por ejemplo, las fechas no pueden contener valores como el 30 de febrero en ningún caso

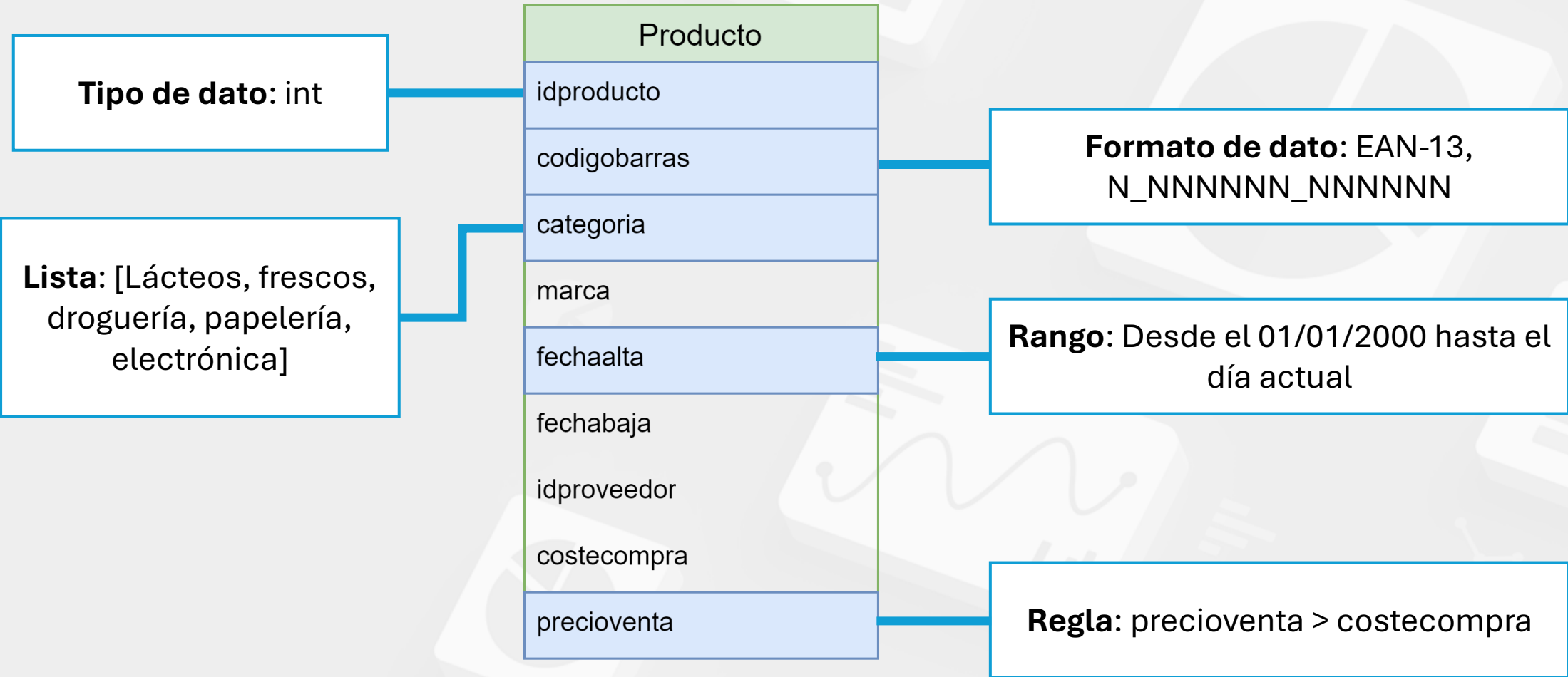
Producto
idproducto
codigobarras
categoria
marca
fechaalta
fechabaja
idproveedor
costecompra
precioventa

Dominios, definiciones

Existen múltiples formas de definir dominios:

- **Tipo de dato:** Limita el tipo de dato (data, int, char(10)...)
- **Formato de dato:** Define patrones y máscaras, como que un teléfono debe tener 9 dígitos o un CIF son 7 dígitos precedidos de una letra y seguidos de un carácter de control calculable
- **Lista:** Listado cerrado de valores válidos
- **Rango:** Describe rangos válidos como números mayores de un valor, fechas anteriores o posteriores hasta 3 meses al día de hoy
- **Basado en reglas:** Incluye reglas de comparación entre atributos, como que una fecha debe ser posterior a otra o que un importe como el precio debe ser superior al coste

Dominios, ejemplos



Claves

Una clave o identificador es un conjunto de uno o más atributos que definen de manera única una instancia de una entidad. Se pueden clasificar y tipificar de las siguientes formas:

- Por tipo de construcción
- Por tipo de función

Producto
idproducto
codigobarras
categoria
marca
fechaalta
fechabaja
idproveedor
costecompra
precioventa

Claves por tipo de construcción

- Una **clave simple es un atributo** que identifica de forma exclusiva una instancia de entidad
- Una **clave subrogada o sustituta es una clave simple creada artificialmente** para la entidad, por ejemplo, un autonumérico. No deben ser visibles para los usuarios
- **Un clave compuesta es un conjunto de atributos** que identifica de forma exclusiva una instancia de entidad
- Una clave compuesta puede estar formada por una clave simple y uno o más atributos no clave

Cliente	
idcliente	Subrogada
dni	Simple
nombre	Compuesta
telefono	
direccion	
email	

Claves por tipo de función

- Una **superclave** es cualquier conjunto de atributos que identifican de manera única una instancia de entidad
- Una **clave candidata** es un conjunto mínimo de uno o más atributos que identifica la instancia de la entidad a la que pertenece
- Una entidad puede tener múltiples claves candidatas
- La clave seleccionada es la **clave primaria (PK)**
- Las claves descartadas se denominan **claves alternas**

Cliente	
idcliente	Superclave / Primaria
dni	Alternativa / Candidata
nombre	Alternativa / Candidata
telefono	
direccion	
email	

Relación

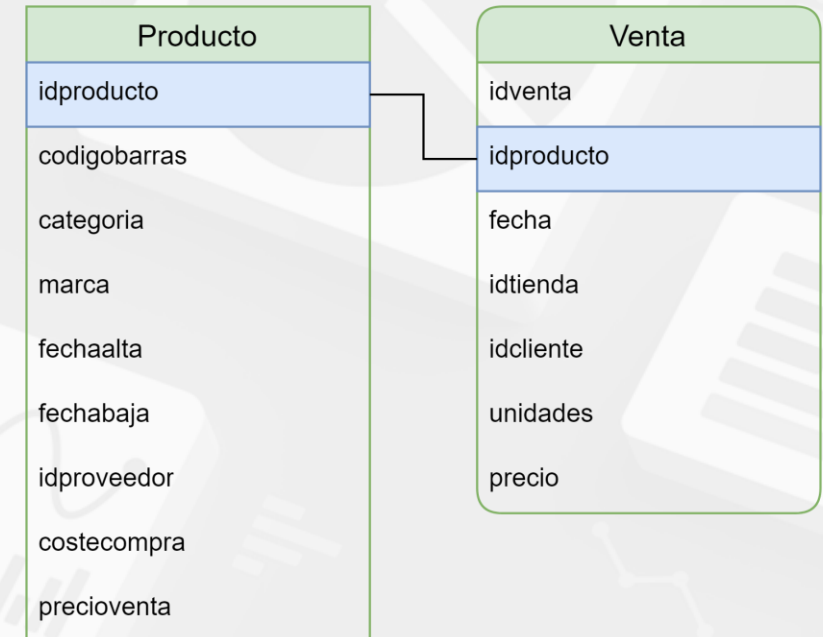
Es una asociación entre entidades

- Captura interacciones de alto nivel entre entidades conceptuales
- Identifica las interacciones detalladas entre entidades lógicas
- Indica las limitaciones entre entidades físicas



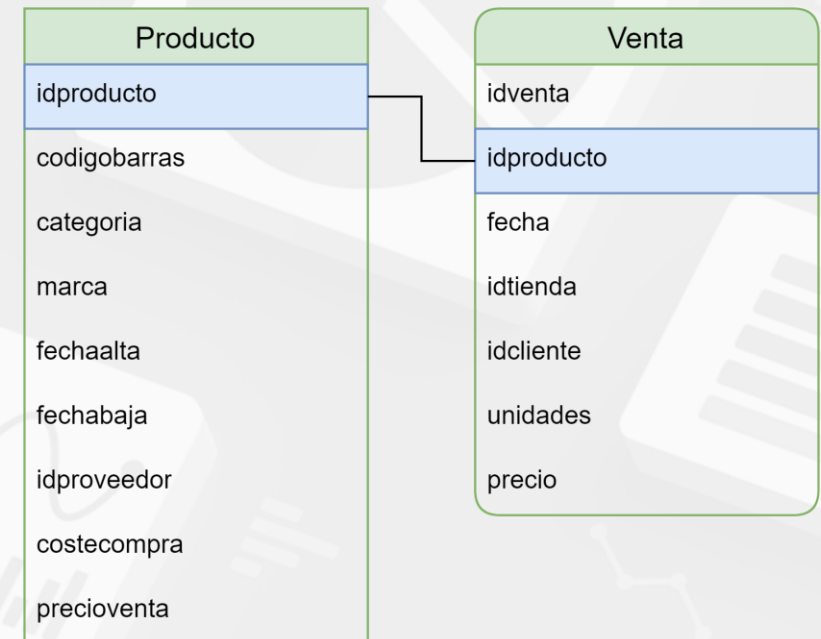
Relaciones identificativas

- Una **relación identificativa** contiene la clave primaria de la entidad padre
- Una **entidad independiente** contiene sólo atributos propios
- Una **entidad dependiente** es aquella cuya clave primaria contiene al menos un atributo de otra entidad. Se suele representar con los bordes redondeados
- Una entidad dependiente tiene al menos una relación identificativa



Claves externas

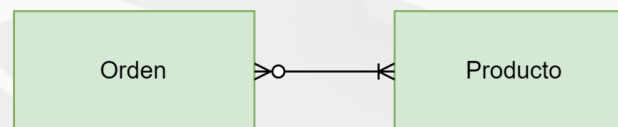
- **Clave externa (Foreign Key, FK):** Campo que contiene el valor de la PK de otra tabla. Se utiliza para generar relaciones y que se puedan realizar consultas a múltiples tablas
- Se utilizan en modelos físicos y lógicos
- Forman parte de **relaciones identificativas**



Cardinalidad de la relación

En la relación entre dos entidades la cardinalidad indica cuántas instancias de cada entidad participan a cada lado de la relación. Los valores que puede tener son 0, 1 o ‘muchos’

Cardinalidad	Uno	Muchos	Uno y sólo uno	Cero o uno	Uno o muchos	Cero o muchos
Representación	1	n	1..1	0..1	1..n	0..n
Símbolo	—+	—>	—++	—o+	—>	—o>

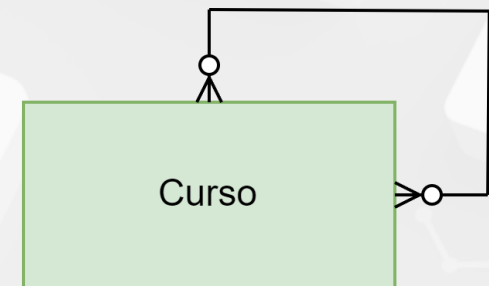
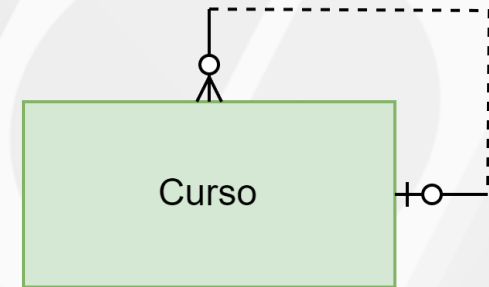


Valencia de la relación

Se llama valencia de una relación al **número de entidades** que forman parte de la misma

- **Relación unaria (recursiva)**

- Representa la relación de una entidad consigo misma
- Si la relación es algo a muchos representa una jerarquía
- Si la relación es de muchos a muchos representa una red o grafo
- En el primer ejemplo un curso puede tener 0 o n prerrequisitos, pero no puede a su vez ser requisito de otro
- En el segundo ocurre lo mismo pero el curso si que puede ser requisito



Valencias binarias y ternarias

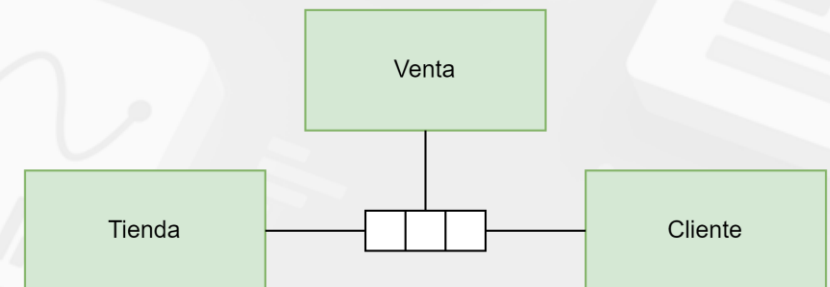
Valencia binaria

- Se da cuando una relación involucra 2 entidades
- Es, con mucho, la más común de las relaciones



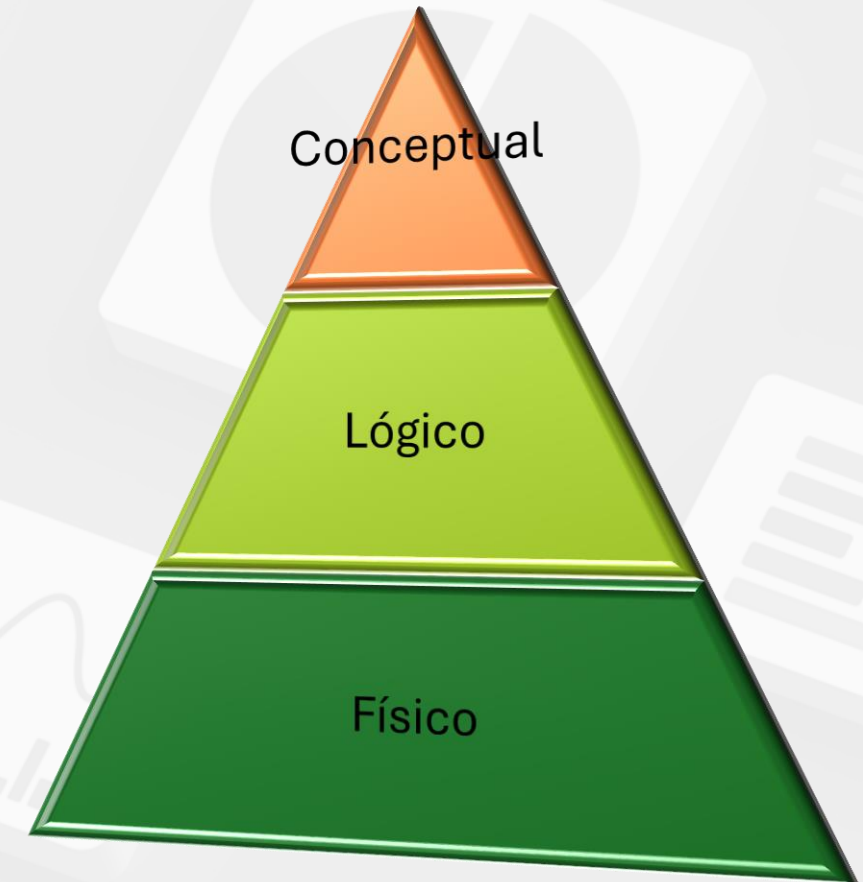
Valencia ternaria

- Se da cuando una relación involucra 3 entidades
- Muy utilizada en el modelado por hechos



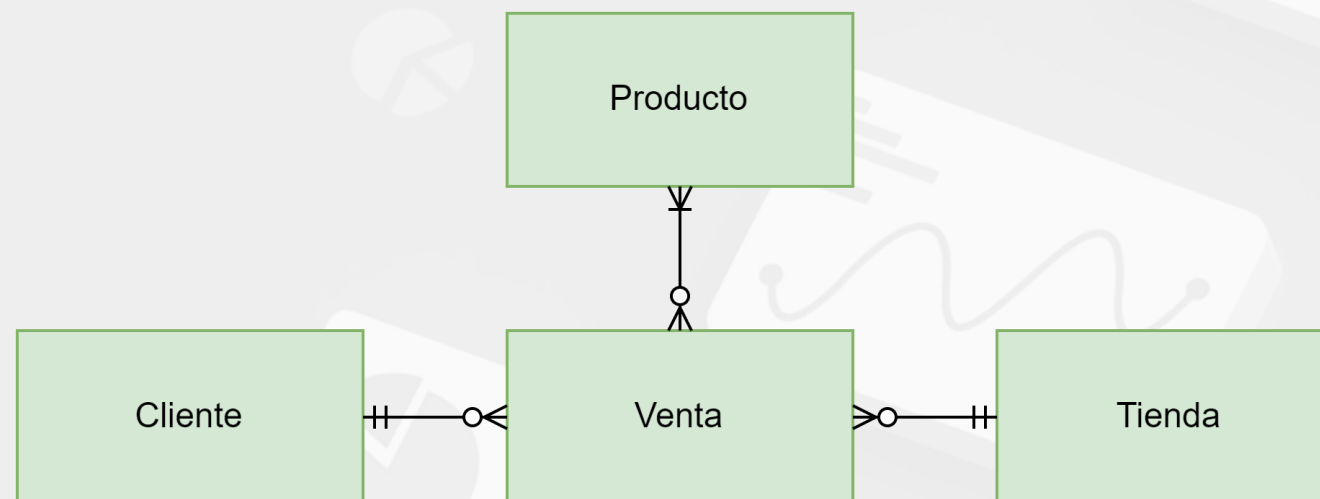
Niveles de detalle del modelo de datos

- **Conceptual:** Representa requisitos de alto nivel como una colección de conceptos relacionados
- **Lógico:** Representación detallada de los requerimientos de datos, habitualmente en apoyo de un contexto específico
- **Físico:** Solución técnica detallada y adaptada al conjunto de hardware y software utilizado



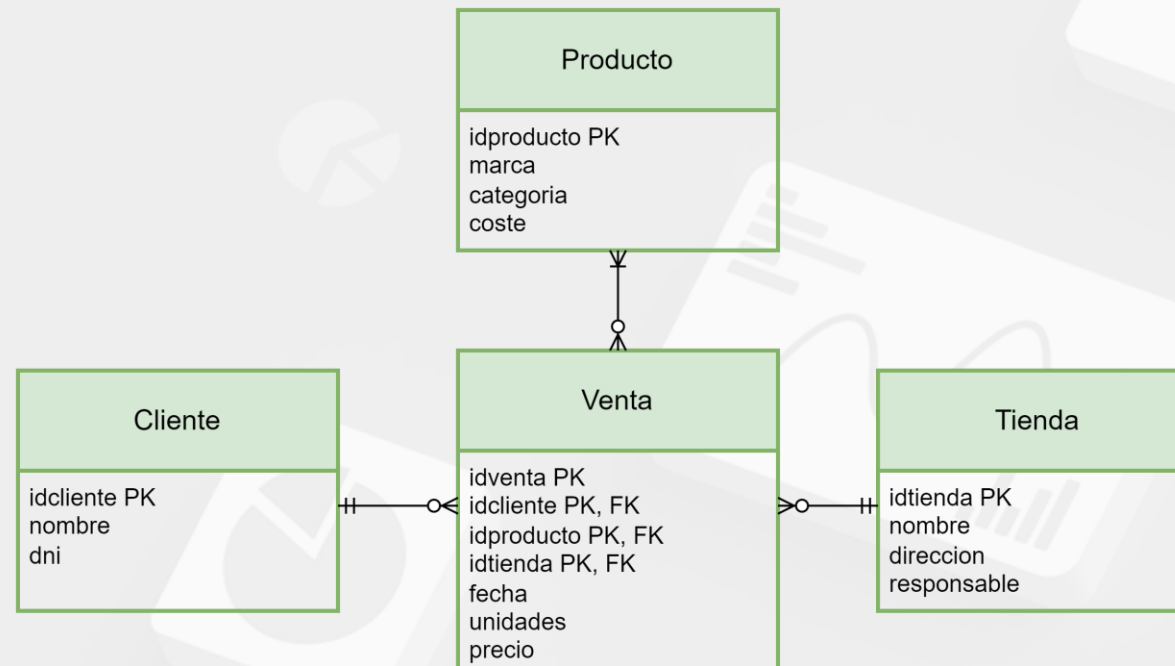
Modelo conceptual (CDM)

- Contiene sólo entidades de negocio básico, su descripción y las relaciones



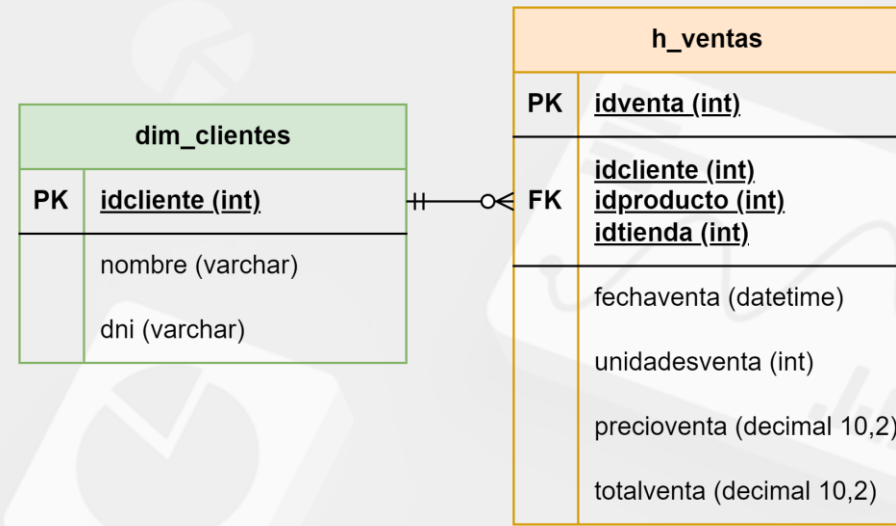
Modelo lógico (LDM)

- Se suelen crear extendiendo un modelo conceptual
- Son independientes de la implementación hardware y software



Modelo físico (PDM)

- Partiendo del modelo lógico, se adapta y amplía siguiendo la implantación de hardware y software



Variantes PDM y técnicas

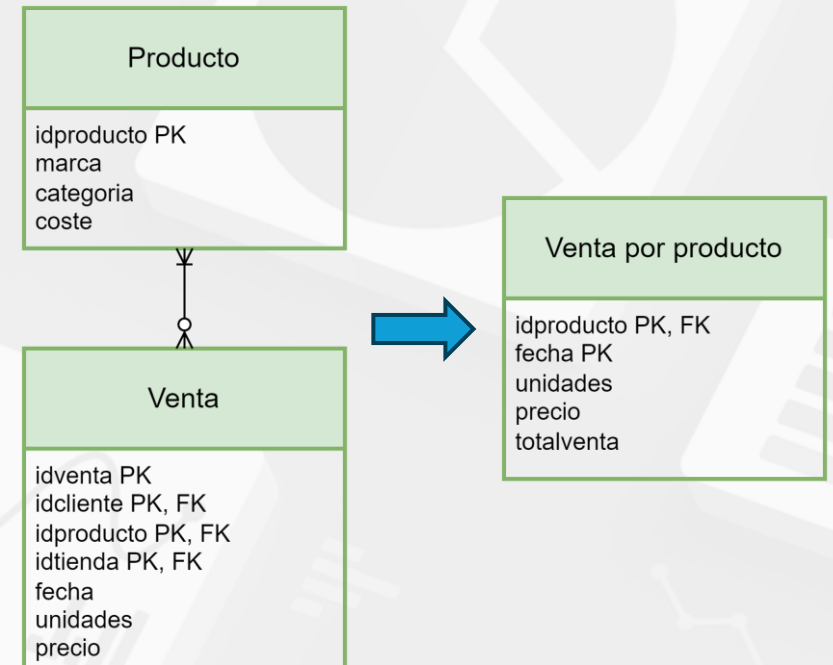
- **Modelo canónico:** Describe la estructura de datos compartida entre sistemas como paquetes o mensajes. Indica qué estructuras utilizar con Web Services, buses empresariales e integración de aplicaciones
- **Vistas:** Son tablas virtuales
 - Ayudan a simplificar consultas, controlar accesos y renombrar columnas
 - El motor de BBDD lanza una consulta SQL en cada acceso
 - Una vista materializada o instanciada almacena y refresca sus datos físicamente de forma periódica
- **Particionamiento:** Proceso de dividir una tabla en partes más pequeñas y manejables, en las BBDD hay utilidades que permiten aplicarlas de forma automática
 - Vertical: División por columnas, se crean tablas con un subconjunto de columnas
 - Horizontal: Se filtran los datos por valores de una o varias columnas

Desnormalización

Transformación deliberada creando tablas con estructuras de datos redundantes o duplicados. Se ponen atributos en varios lugares de forma intencionada. Se hace para mejorar el rendimiento:

- Combinar datos de varias tablas para evitar joins
- Crear copias de datos previamente filtrados
- Precálculo y almacenamiento de datos

Se introduce un riesgo de error por duplicados y hace más complejo el mantenimiento para mantener los datos actualizados



Normalización

Proceso de aplicación de reglas para **crear estructuras de datos estables** manteniendo cada atributo en un lugar único eliminando redundancia e inconsistencias.

Cada nivel de normalización no tiene porqué incluir al anterior (aunque facilita su implementación) y **soluciona problemas adicionales**.

Se considera que los datos están **normalizados cuando están en 3NF**. Por encima de la 3NF las formas se consideran ejercicios académicos con una aplicación muy excepcional.

Normalización, formas normales

- **Primera forma normal (1NF):** Asegura que cada entidad tenga una clave primaria válida, y que cada atributo dependa de la clave primaria. Elimina grupos repetidos y asegura que cada atributo sea atómico (sin multi valor). 1NF incluye la resolución de relaciones de muchos a muchos con una entidad adicional a menudo llamada entidad asociativa o puente.
- **Segunda forma normal (2NF):** Asegura que cada entidad tiene una clave primaria mínima y que cada atributo depende de la clave primaria completa.
- **Tercera forma normal (3NF):** Asegura que cada entidad no tiene claves primarias ocultas y que cada atributo no depende de ningún atributo fuera de la clave. “La clave, la clave completa y nada más que la clave”.
- **Forma normal de Boyce Codd (BCNF):** Resuelve la superposición de claves candidatas compuestas. Clave superpuesta significa que hay reglas de negocio ocultas entre las claves.
- **Cuarta forma normal (4NF):** Resuelve todas las relaciones de muchos a muchos a muchos (y más allá) en pares hasta que no pueden ser desglosadas en bloques más pequeños.
- **Quinta forma normal (5NF):** Resuelve las dependencias entre entidades en pares básicos, y todas las dependencias unidas utilizan partes de las claves primarias.

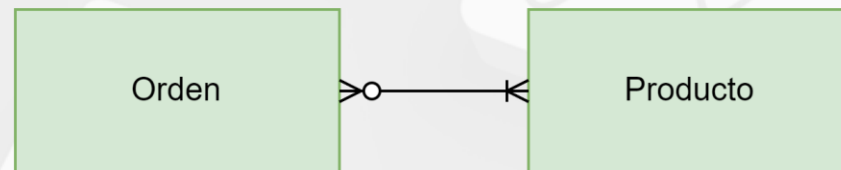
Esquemas de modelado de datos

- El esquema y la notación depende de la tecnología utilizada

Esquema	Ejemplo de notaciones
Relacional	Information Engineering (IE) Integration Definition for Information Modeling (IDEF1X) Barker Chen
Dimensional	Dimensional
Orientado a objetos	Unified Modeling Language (UML)
Basado en hechos	Object Role Modeling (ORM / ORM2) Fully Communication Oriented Modeling (FCO-IM)
Basado en tiempo	Data Vault Modelado tipo ancla
NoSQL	Documento Columna Grafo Clave-Valor

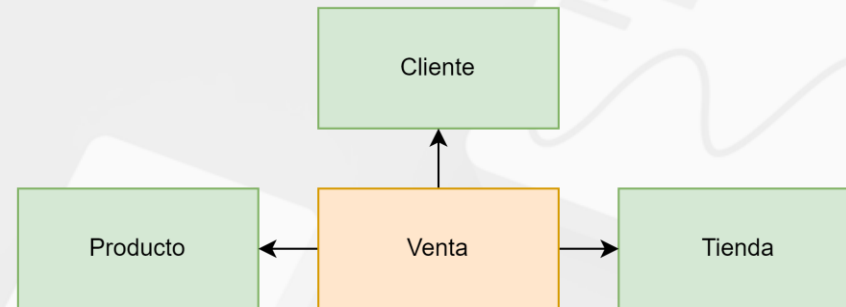
Relacional

- El objetivo de diseño es conseguir una expresión exacta de los datos de negocio y tener un hecho en un solo lugar eliminando la redundancia
- Es ideal para el diseño de sistemas operacionales
- Notación más habitual es Information Engineering (IE):



Modelo dimensional

- Los datos se estructuran para **optimizar la consulta y el análisis** de grandes cantidades de datos
- Recogen las **preguntas de negocio** sobre procesos concretos
- Se basan en dimensiones y medidas o hechos



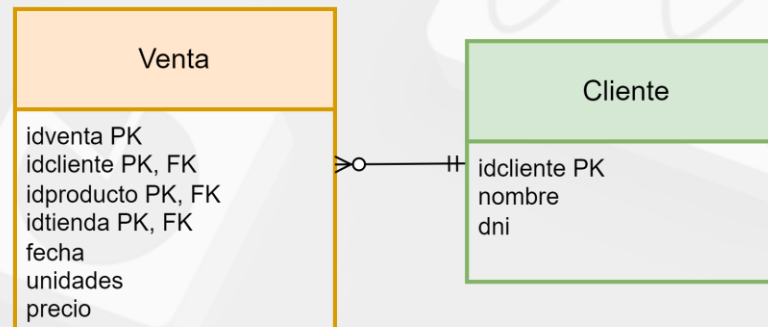
Tablas del modelo dimensional

Tablas de hechos (fact tables)

- Las filas corresponden a **medidas y son numéricas**, tales como importes, cantidades o conteos
- Suelen representar en torno al 90% del espacio ocupado

Tablas de dimensiones

- Representan los **objetos importantes de la empresa** y contienen sobre todo descripciones textuales
- Suelen ocupar en torno al 10% del espacio



Dimensiones SCD

- SCD (Slowly Changing Dimensions) o lentamente cambiantes
- Se trata de una técnica que permite gestionar **cambios en valores de una dimensión**:
 - **Tipo 1**: el dato nuevo sobrescribe el antiguo
 - **Tipo 2**: Se crea un registro para cada valor, se gestiona mediante campos de fecha y flags
 - **Tipo 3**: Cada nuevo valor se mete en una columna

Orientado a objetos y basado en hechos

Orientado a objetos

- Utiliza UML mediante objetos tipo clase
- Es similar a un modelo de entidad-relación pero añade los métodos
- Se basa en el código en el que se va a implementar

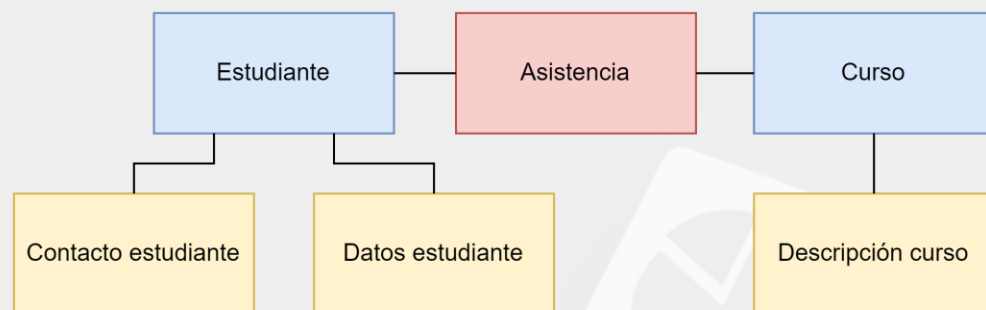
Basado en hechos

- Se basa en lenguajes de modelado conceptual
- Tratan de analizar la verbalización del lenguaje natural
- Object Role Modeling (ORM/ORM2) se basa en una versión restringida de lenguaje natural no ambigua
- Fully Communication Oriented Modeling (FCO-IM) es otro lenguaje similar

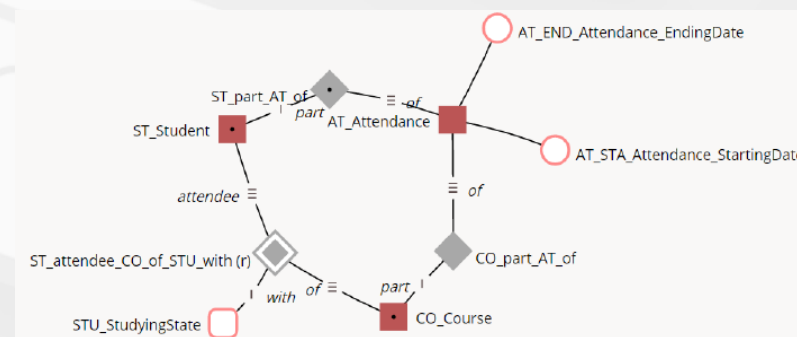
Basado en tiempo

Los valores deben estar asociados en orden cronológico y con valores de tiempo específico

- Data Vault: Se trata de un híbrido de 3NF y esquema en estrella. Contiene hubs (azul), enlaces (rojo) y satélites (amarillo)



- Modelado tipo ancla



NoSQL

- Tecnologías no relacionales, originalmente eran No SQL, hoy se dice Not Only SQL
- Sus tipos están basados en la forma de almacenar los datos:
 - **Documento:** Almacenan estructuras completas llamadas documentos
 - **Clave-Valor:** Guardan los datos en dos columnas, una de clave y otra de valor que puede contener casi cualquier tipo de información
 - **Orientado a columna:** Puede parecer similar a una base de datos relacional, pero aparte de tipos de datos complejos los registros de una tabla pueden tener estructuras diferentes
 - **Grafo:** Se basan en nodos y sus conexiones. Son ideales para analizar relaciones



Plan de modelado de datos

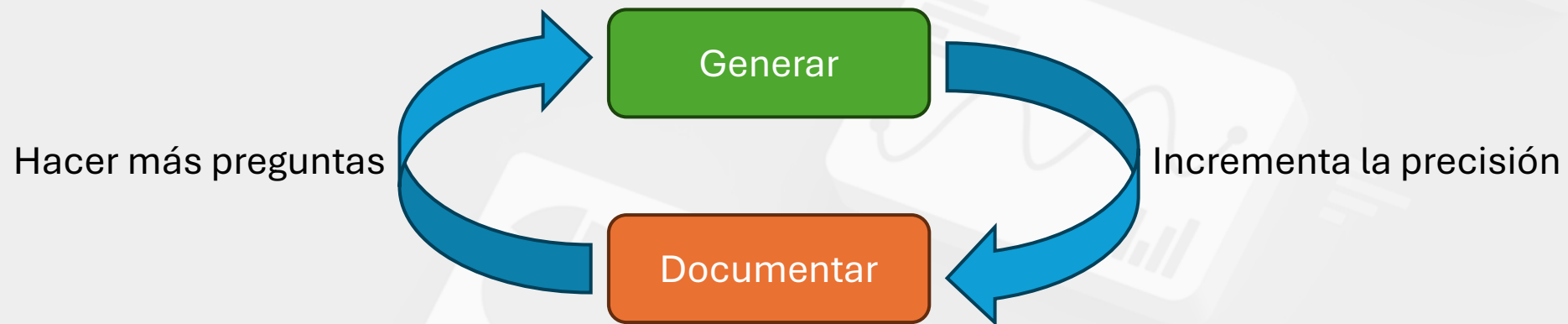
Debe contener tareas como **evaluar requerimientos organizacionales**, creando estándares y definiendo el almacenamiento.

Puede incluir los entregables siguientes:

- **Diagrama:** Forma visual que recoge los requerimientos de forma precisa. Se define por nivel de detalle, tipo de esquema y notación
- **Definiciones:** Descripción de las entidades, atributos y relaciones
- **Problemas y cuestiones pendientes:** Listado de cuestiones pendientes de resolución, habitualmente por falta de contexto de negocio. Se resolverán en iteraciones posteriores y durante el propio proceso de modelado
- **Linaje:** Mapeos entre origen y destino. Relación entre componentes de los modelos conceptual, lógico y físico

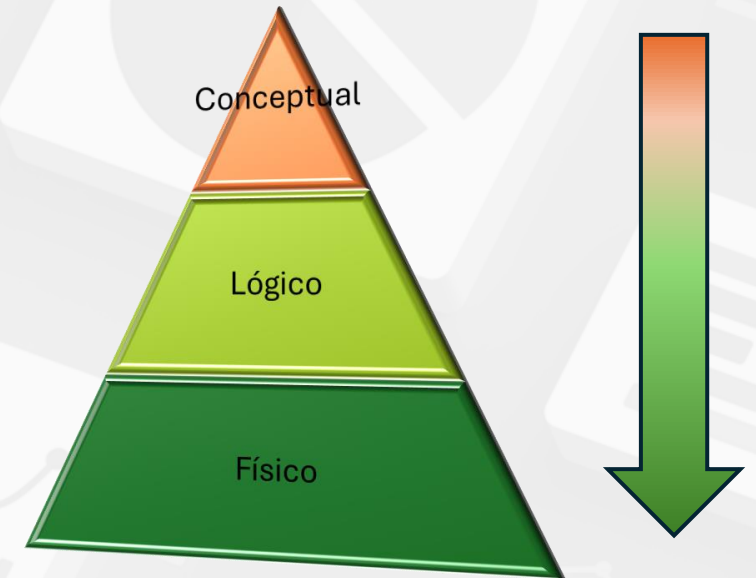
Construir el modelo de datos

- La construcción de un modelo es un proceso iterativo
- Se prepara una versión y se coteja con el negocio
- Se revisa el modelo, se generan nuevas preguntas y el ciclo comienza de nuevo

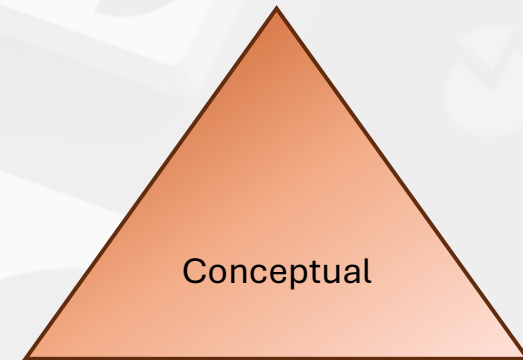


Construir el modelo, ingeniería directa

- Es el proceso de documentar partiendo de los requerimientos
- Se comienza con el CDM
- El LDM se crea para documentar la solución de negocio
- Finalmente se trabaja con el PDM para definir la parte técnica



Ingeniería directa, modelo conceptual



- **Seleccionar esquema**
- **Elegir notación**
- **CDM inicial:** Diseñar pensando en un grupo de usuarios
 - Recoger conceptos (sustantivos) de alto nivel
 - Identificar las actividades (verbos) que se conectan a estos conceptos
- **Terminología de la organización:** Asegurar el alineamiento con la nomenclatura utilizada
- **Aprobación:** Verificar buenas prácticas y alineamiento con los requisitos, después conseguir una aprobación explícita

Ingeniería directa, modelo lógico



- **Analizar requerimientos de información:** Identificar las necesidades de negocio en el contexto de procesos de negocio
- **Analizar documentación existente:** Revisar artefactos de datos existentes, validar su alineamiento con los requisitos
- **Agregar entidades asociativas:** Romper las relaciones muchos a muchos del modelo
- **Añadir atributos:** Crear los atributos que sean necesarios para cada entidad. Definir el dominio para cada uno de ellos e identificar cuáles van a funcionar como claves

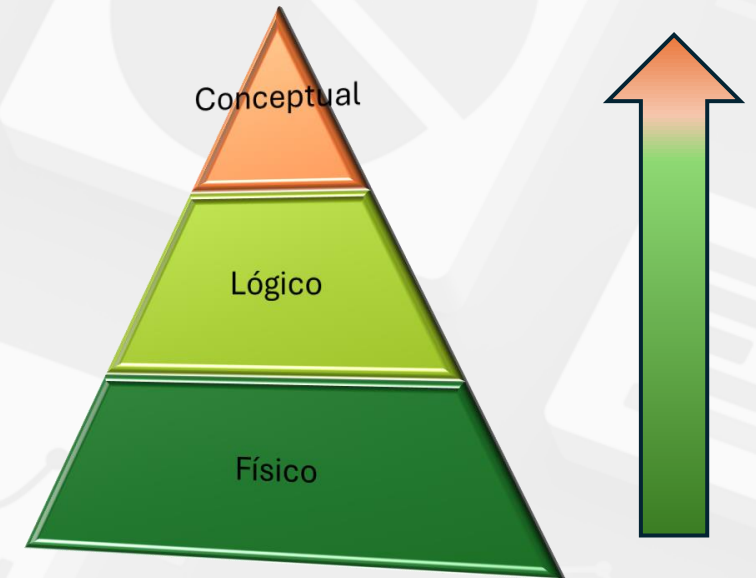
Ingeniería directa, modelo físico



- **Resolver abstracciones:** Crear los objetos necesarios para soportar súper tipos y subtipos
- **Añadir detalle de atributos:** Indicar nombres técnicos, esquema, tipo de datos asociados a cada entidad y atributo
- **Añadir objetos de datos de referencia:** Crear objetos para soportar datos de referencia, definir reglas de mantenimiento
- **Asignar claves subrogadas:** Incluir el cálculo de las claves manteniendo la clave natural como clave alterna
- **Mejorar el rendimiento:** Utilizar técnicas como la desnormalización o la creación de índices y particiones para mejorar el rendimiento de las consultas
- **Crear vistas:** Utilizar vistas para controlar el acceso a datos y simplificar la ejecución de consultas frecuentes

Construir el modelo, ingeniería inversa

- Es el proceso de documentar una base de datos existente
- Se comienza con el PDM
- El LDM se crea o actualiza partiendo del PDM
- Finalmente se trabaja con el CDM tomando el LDM como base



Revisión y mantenimiento del modelo de datos

- Los modelos requieren de un **control de calidad** e implementar medidas de mejora continua
- Se recomienda implantar técnicas como el Data Model Scorecard para evaluar la calidad de los modelos
- Los modelos a diferente nivel de detalle se deben **mantener alineados**
- Tras un desarrollo es recomendable aplicar ingeniería inversa para **asegurar la coherencia**

Herramientas, modelado y linaje

- **Modelado de datos:** Permiten automatizar muchas tareas de modelado, proporcionan funcionalidad de dibujo y paletas de fácil uso. Pueden generar lenguaje DDL (Database Definition Language) para pasar del modelo lógico al físico. En ciertos casos permiten validar estándares como los de nomenclatura.
- **Linaje:** Contiene información sobre el código de carga y los modelos de datos. Ayuda a investigar causas de problemas y determinar la fiabilidad de los datos. Permite generar análisis de impacto.

Herramientas, perfilado y repositorio de metadatos

- **Perfilado de datos:** Ayuda a explorar el contenido de los datos, validarlo e identificar problemas de calidad. Permiten también levantar errores en las definiciones de las entidades y atributos comparando los datos con las definiciones creadas.
- **Repositorio de metadatos:** Describe los datos en términos de negocio e incluye otra información para usarlos. Debe disponer de interfaces sencillos para la explotación y uso de los metadatos.

Patrones y modelos de datos de industria

- **Patrones de modelo de datos:** Estructuras de modelado y patrones de integración reutilizables aplicables para solventar ciertas situaciones. Por ejemplo, permiten resolver relaciones de muchos a muchos o ayudan a construir jerarquías auto referenciadas.
- **Modelos de datos de industria:** Son modelos de datos preconstruidos para todo un sector, como salud o banca. Son de amplio alcance, muy detallados y se deben adaptar y ajustar a la realidad de cada organización. Pueden adquirirse a través de proveedores o grupos industriales. Varios ejemplos: ARTS (comercio minorista), SID (telecomunicaciones) o ACORD (seguros)

Buenas prácticas, nomenclatura

- Debe **definirse de forma conjunta** entre arquitectos, analistas y administradores, pero sin perder de vista los requisitos del negocio
- Comprende los nombres de atributos, campos, tablas, vistas y demás objetos de los modelos
- La nomenclatura debe ser **compartida** en toda la organización
- Requisitos generales:
 - Acorde a reglamentos de IT
 - Nombres lógicos con sentido de negocio
 - Se deben evitar las abreviaciones salvo las más comunes acordadas
 - Se recomienda utilizar palabras completas con separación con espacios para los nombres lógicos y barras bajas para los físicos
 - Se debe tener en cuenta y aprovechar la longitud máxima soportada por el motor de base de datos utilizado
 - No se debe incluir el entorno (desarrollo, pruebas...) en el nombre de los objetos

Buenas prácticas, diseño de base de datos

Para la construcción de la base de datos se recomienda seguir los siguientes principios (acrónimo PRISM):

- **Performance:** garantizar el buen rendimiento y el acceso fácil de los usuarios
- **Reusability:** evitar acoplamiento con una aplicación, diseñar pensando en un uso más extendido de los datos
- **Integrity:** asegurar que los datos son válidos y útiles para el negocio. Acercar las restricciones de calidad e integridad lo más cerca del origen
- **Security:** datos exactos y válidos, pero sólo para los usuarios autorizados. Vigilar los permisos y accesos
- **Maintainability:** evitar que el coste de evolucionar o cambiar los procesos sea superior al de su desarrollo. Diseñar sin perder de vista el soporte futuro

Gestión de la calidad del modelado y diseño

Para asegurar la calidad del modelo se debe definir un **estándar de diseño y modelado** que contenga por lo menos lo siguiente:

- Lista y descripción de entregables
- Reglas de nomenclatura
- Lista y descripción de roles y responsabilidades
- Catálogo de metadatos
- Buenas prácticas de diseño y modelado
- Directrices para revisiones de diseño
- Método de control de versiones
- Compendio de prácticas no recomendables

Para mantener la calidad se deben realizar **reuniones de revisión periódicas** para validar que las modificaciones introducidas al modelo están alineadas tanto con las directrices como con los objetivos de negocio.

Métricas de modelado de datos

Se deben definir métricas para cuantificar cómo de bien funciona un modelo. Se propone una definición de “The Data Model Scorecard”:

#	Categoría
1	¿Cómo de bien captura el modelo los requerimientos?
2	¿Cómo de completo es el modelo?
3	¿Cómo de bien se ajusta el modelo a su esquema?
4	¿Cómo de sólido estructuralmente es el modelo?
5	¿Cómo de bien aprovecha el modelo las estructuras genéricas?

#	Categoría
6	¿Cómo de bien se siguen los estándares de nomenclatura?
7	¿Cómo de bien se ha organizado el modelo para facilitar la legibilidad?
8	¿Cómo de buenas son las definiciones?
9	¿Cómo de coherente es el modelo con la organización?
10	¿Cómo de bien se ajustan los metadatos a los datos?

Preguntas de ejemplo CDMP

11%

A deliverable in the data modelling and design context diagram is the logical data model.

- A. True
- B. False

A deliverable in the data modelling and design context diagram is the logical data model.

- A. True
- B. False

SDLC stands for:

- A. System development leverage cycle
- B. System design lifecycle
- C. System and design long cycle
- D. System development lifecycle

SDLC stands for:

- A. System development leverage cycle
- B. System design lifecycle
- C. System and design long cycle
- D. System development lifecycle

Business activity information is one of the types of data that can be modelled.

- A. True
- B. False

Business activity information is one of the types of data that can be modelled.

- A. True
- B. False

The number of entities in a relationship is the arity of the relationship. The most common are (multiple choice):

- A. Unary
- B. Binary
- C. Trinary
- D. Ternary

The number of entities in a relationship is the arity of the relationship. The most common are (multiple choice):

- A. Unary
- B. Binary
- C. Trinary
- D. Ternary

What type of key is used in physical and sometimes logical relational data modelling schemes to represent a relationship?

- A. Primary key
- B. Foreign key
- C. Network key
- D. Applications key
- E. Door key
- F. All of the above

What type of key is used in physical and sometimes logical relational data modelling schemes to represent a relationship?

- A. Primary key
- B. Foreign key
- C. Network key
- D. Applications key
- E. Door key
- F. All of the above

Please select valid modelling schemes or notations (multiple choice):

- A. NoSQL
- B. Dimensional
- C. Relational
- D. Object-orientated
- E. Fact-based
- F. Matrix-based

Please select valid modelling schemes or notations (multiple choice):

- A. NoSQL
- B. Dimensional
- C. Relational
- D. Object-orientated
- E. Fact-based
- F. Matrix-based

¡Gracias!



GOLD SPONSORS



SILVER SPONSORS



Entidades académicas

