Sección 02: Un lenguaje común

Vamos a hablar de

* ¿Que son los datos?
* ¿Qué es una base de datos?
* ¿Qué es un modelo de datos?

Dato: Representación simbólica de un atributo o variable cuantitativa o cualitativa.

Simbólica porque es un dibujo, letras, figuras, números.

Cuantitativa: Cuanto es, 1Kg, 3m, 10s.

Cualitativa: Como es, redonda, verde, dulce.

Describe una entidad, un hecho, un momento de la vida real a través de un atributo de estos.

Un dato solo NO tiene información.

Ej: El numero 3 es concreto, conocido, pero no es información ni conocimiento, no sabemos si son 3 naranjas o 3 pelotas.

Metadatos:

Podemos agregarle información, decir que son 3 decimales, 3% de descuento.

Al dato le hemos agregado algo: decimales, %. Eso que le hemos agregado se llama “Metadatos o Metadata” y son datos que describen al dato, le dan sentido. Se llaman de forma genérica “Metadata descriptiva” porque son datos que describen al dato.

También hay datos que describen donde está el dato, El sitio donde está el dato lo llamamos “Contenedor”

Planilla Costos

Campo edad

Estante 4

Estos datos que describen al contenedor se llaman “Metadata Estructural”

En general llamamos a ambos “Metadata”

Un Infante puede tener:

Edad

• Sexo

• Estatura

• Peso

• Fecha Nac. TODOS son ATRIBUTOS de la Relación

• Lugar Nac.

• Padre

• Madre

• Dirección

• Hermanos

• Alergias

Esa agrupación de datos es la que vamos a llamar una “Entidad” o “Relación” y así generamos información.

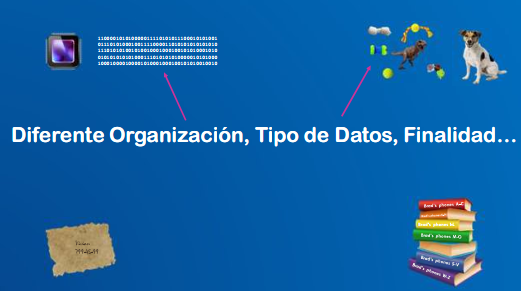
Resumiendo: Datos son

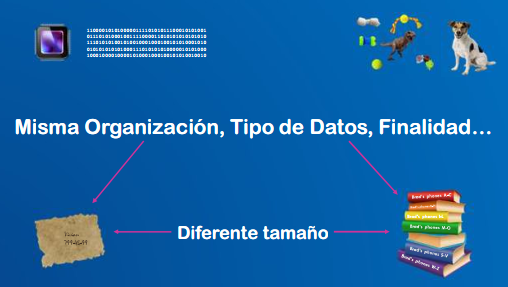
* Atributos que describen Entidades, Hechos y/o Momentos del mundo real.
* Hay datos que describen a los Datos (Metadata).
* Un Dato aislado no tiene sentido. Es irrelevante, incomprensible.
* Aunque es algo concreto, los Datos no son información,

¿Qué es una Base de Datos?

* Colección de información organizada de tal modo que sea fácilmente accesible, gestionada y actualizada.
* Conjunto de datos pertenecientes a un mismo contexto, almacenado sistemáticamente para su posterior uso.
* Conjunto de datos informativos organizados en un mismo contexto para su uso y vinculación.

Vemos que se repiten palabras como: Colección, organizada, accesible, gestionada, actualizada, mismo contexto, sistemáticamente.





La clasificación puede ser:

Por su tasa de modificaciones:

* Libro -> Estático - > Solo lectura
* Tabla Excel -> Dinámica -> ingreso y egreso de datos -> cambios

Por su alcance:

* Personales
* Locales
* Departamentales
* Empresariales
* Regionales
* Mundiales

Por su uso:

* Musicales
* Geográficas
* Juegos
* Bibliográficas
* Científicas

Una base de datos es:

* Colección organizada de datos pertenecientes a un mismo contexto
* La organización y el contexto dependen del “dueño”, el alcance o el manejo de la base de datos.
* Los datos deben estar disponibles y ser gestionables
* Su clasificación es muy subjetiva.

Que NO ES una base de datos:

* SQL es un lenguaje, no es una base de datos
* No es un software ni un programa

¿Qué es un modelo de Datos?

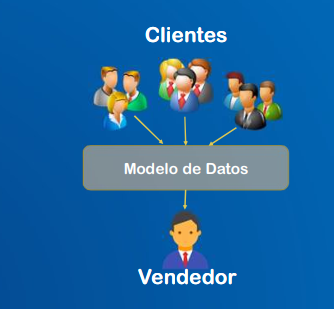
Definición: Es un lenguaje de modelado para la definición de esquemas. Por esquema vamos a entender una representación gráfica o simbólica.

Como interactúan los datos entre si y como se relacionan con el mundo externo.

Se aplican en muchas partes, pero nos vamos a enfocar en la Informática y el desarrollo de bases de datos.

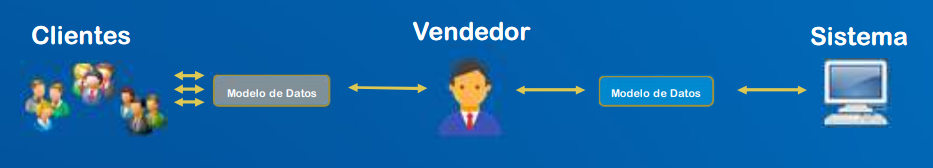
Si tenemos un vendedor y un grupo de clientes puede darse el caso que cada cliente tenga su base de datos y el vendedor también tenga la suya.

Es necesario que entre ellos hablen un mismo idioma

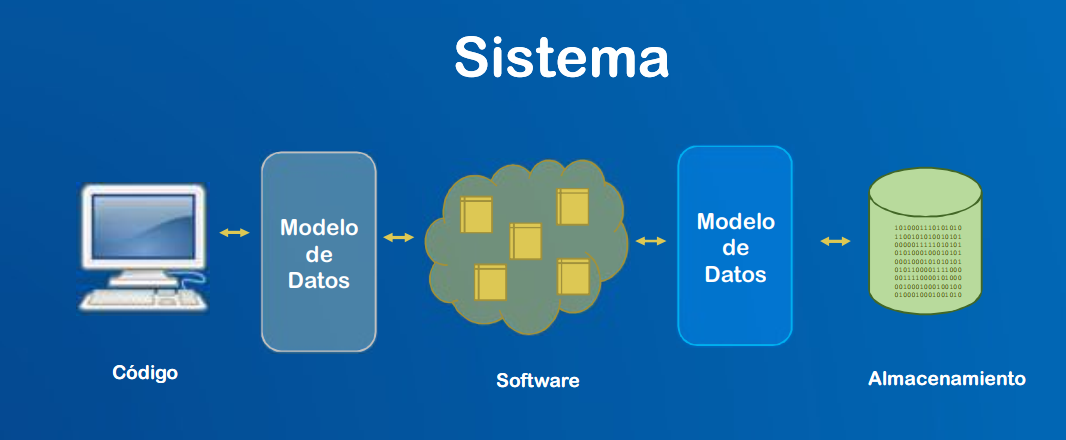


Es necesario que entre ellos se establezca un modelo de datos.

A su vez el vendedor se va a comunicar con un sistema a través de un modelo de datos:



En un sistema también se establece un modelo de datos:



Características de un modelo de datos:

* Un lenguaje común ¿Como vamos a hacer para entendernos?
* Definir entidades ¿De qué y de quienes estamos hablando aquí?
* Reglas del negocio ¿Cómo se hace esto? ¿Qué pasa aquí?

Lenguaje común: Las matemáticas, un idioma

Definir entidades: Clientes y vendedor, vemos que ambos son sustantivos y un sustantivo tiene existencia real.

En una venta se generan varias entidades como por ejemplo productos y facturas.



¡Todos son sustantivos!

Reglas del negocio:

* ¿Cómo se hace esto?
* ¿Qué pasa aquí?



Si cada factura tiene que tener AL MENOS UN PRODUCTO se identifica con una flecha de punta sólida.

El símbolo infinito señala que no hay limite en la cantidad de productos que pueda tener una factura.

Definimos la cardinalidad que es el numero de elementos de un conjunto.

Flecha de punta llena:

Cardinalidad mínima 1

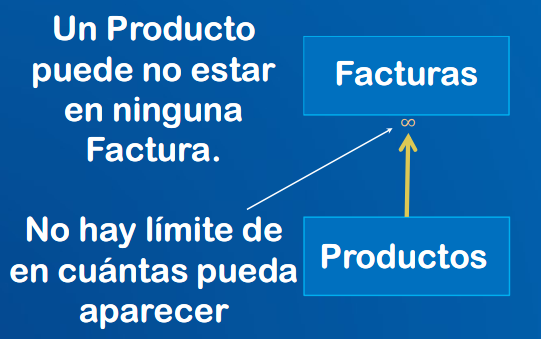
Cardinalidad máxima n

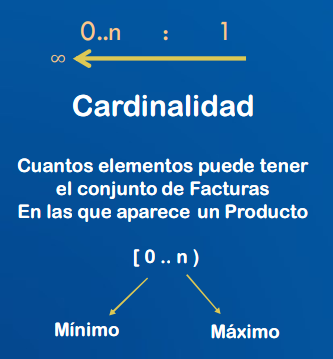
Definimos así la relación entre facturas y productos y ahora tenemos que definir la relación entre productos y facturas que puede ser diferente. Un producto que NO se haya vendido No aparece en ninguna factura.

O sea:

Factura -> 1 a N Productos

Productos -> 0 a N Facturas

La flecha de punta abierta señala que puede haber productos que estén en cero facturas.

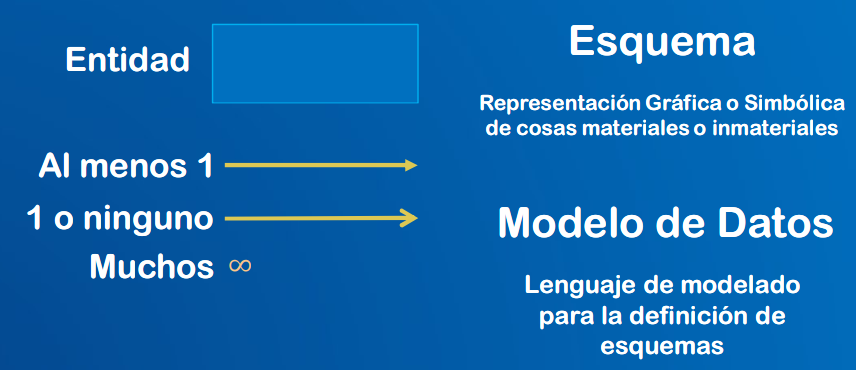
 Flecha de punta abierta

La cardinalidad depende de donde a donde vemos, si de la factura al producto o del producto a la factura

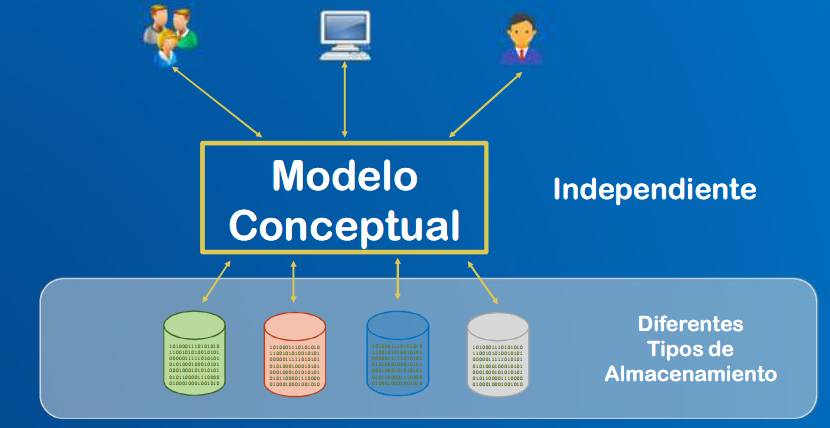


Del producto a la factura De la factura al producto

Hemos establecido una serie de símbolos para ver cuales son las reglas del negocio



Un modelo de datos es un modelo conceptual y tiene que ser claro, no puede ser ambiguo.

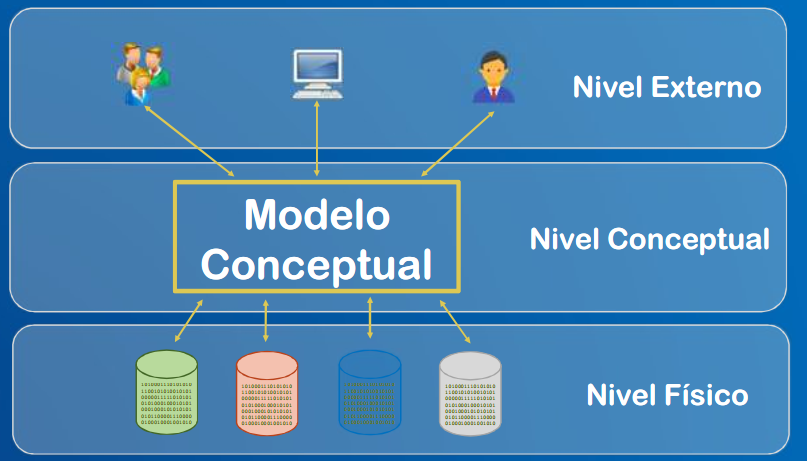


Un modelo conceptual debe ser independiente del dispositivo físico de donde se almacenan los datos.

Un modelo conceptual debe ser

* Flexible
* Estable
* Reutilizable

ANSI define 3 niveles:



Recapitulando, un Modelo de Datos es:

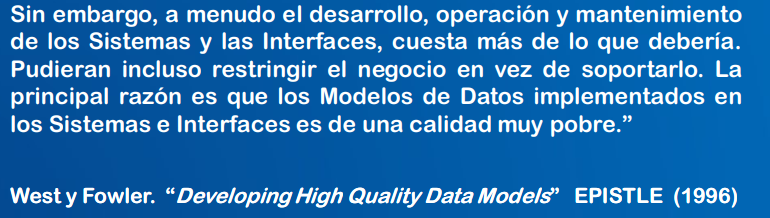
* Es una estructura abstracta
* Documenta, Organiza y estandariza la información
* Define un lenguaje común para entendernos
* Identifica las entidades
* Describe las reglas del negocio (Relaciones)

Un modelo de datos debe ser:

* Claro, no ambiguo
* Independiente de los dispositivos
* Estable, Flexible, reutilizable

ANSI define 3 niveles para los modelos de datos: ( no es la única forma de verlos)

* Nivel externo
* Nivel conceptual
* Nivel físico



**Hasta aquí:**

*Dato: Representación simbólica de un atributo o variable cuantitativa o cualitativa.*

*Resumiendo: Datos son*

* *Atributos que describen Entidades, Hechos y/o Momentos del mundo real.*
* *Hay datos que describen a los Datos (Metadata).*
* *Un Dato aislado no tiene sentido. Es irrelevante, incomprensible.*
* *Aunque es algo concreto, los Datos no son información,*

*¿Qué es una Base de Datos?*

* *Colección de información organizada de tal modo que sea fácilmente accesible, gestionada y actualizada.*
* *Conjunto de datos pertenecientes a un mismo contexto, almacenado sistemáticamente para su posterior uso.*
* *Conjunto de datos informativos organizados en un mismo contexto para su uso y vinculación.*

*Resumiendo:*

* *Colección organizada de datos pertenecientes a un mismo contexto*
* *La organización y el contexto dependen del “dueño”, el alcance o el manejo de la base de datos.*
* *Los datos deben estar disponibles y ser gestionables*
* *Su clasificación es muy subjetiva.*

*¿Qué es un modelo de Datos?*

*Definición: Es un lenguaje de modelado para la definición de esquemas. Por esquema vamos a entender una representación gráfica o simbólica.*

*Recapitulando, un Modelo de Datos es:*

* *Es una estructura abstracta*
* *Documenta, Organiza y estandariza la información*
* *Define un lenguaje común para entendernos*
* *Identifica las entidades*
* *Describe las reglas del negocio (Relaciones)*

Modelos de Bases de Datos:

El modelo más utilizado hasta el día de hoy, el modelo Relacional, es un desarrollo del año 1969 por Edgar Codd.

Vamos a ver los 4 modelos mas comunes de bases de datos:

1. Modelo de archivos planos
2. Modelo jerárquico
3. Modelo de redes
4. Introducción al modelo relacional

1) Modelo de Archivos Planos: No es un modelo de base de datos, pero es la base de como se almacenan los datos en las computadoras. Es como si fuese una hoja de papel donde ponemos de forma libre los datos.



Podemos separar los datos por algún carácter, como un coma.

Poner marcadores <td> para separara los datos o totalmente sin formato, con un texto de una receta de cocina.

La metadata NO está presente dentro del archivo y debe ser conocida por ambas partes para entender el mismo. Esto en el mundo de la informática se llama contrato.

Una característica de los textos planos es que podemos leerlos, ver cuál es el texto

Los archivos binarios son archivos de textos planos y decimos que “No son archivos de texto”, no los podemos leer. Ej: procesador de texto, planilla de cálculo.

Desventajas

* Estructuras poco Organizadas
* Solo sirven para Pequeños Volúmenes de datos
* Carecen de Seguridad
* Muy difícil Relacionarlas (casi imposible)
* Es muy fácil corromper los datos

Ventajas

* Fáciles de Comprimir
* Transferencia de Datos
* Formatos humanamente inteligibles

¡Toda la información en una computadora reside en algún tipo de archivo plano!

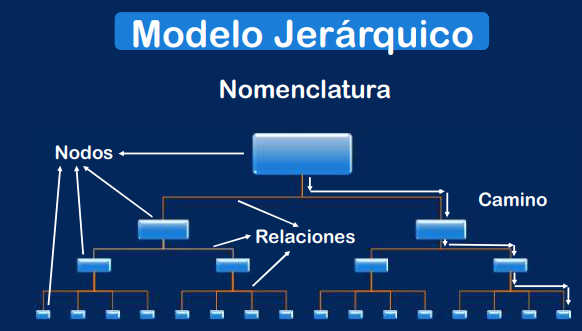
2) Modelo Jerárquico

Fue le modelo que dio inicio a los modelos de bases de datos, los datos están organizados en una estructura jerárquica del tipo árbol invertido.

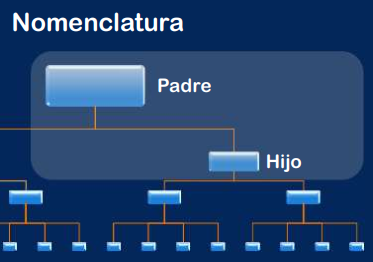
El modelo jerarquico tiene una Raíz, Ramas y Hojas

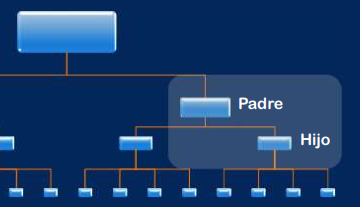


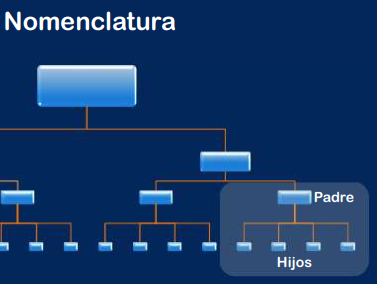
Un conjunto de relaciones es un camino entre un nodo y otro



Dos nodos que se relacionan entre si se denominan Padre e Hijo, a cualquier nivel que se relacionen.



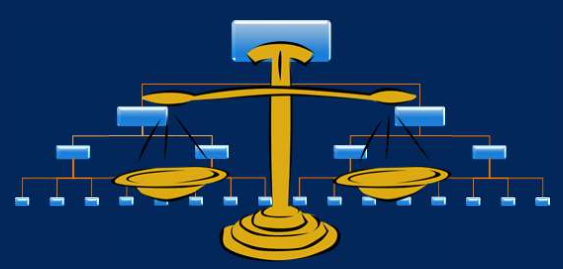




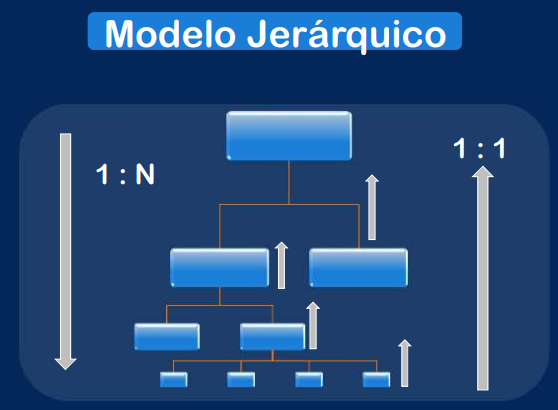
Dos nodos del mismo nivel se denominan hermanos sin importar el nivel



Árbol balanceado: Ambas ramas están balanceadas, puede haber diagramas que no estén balanceados



Si recorremos los nodos de arriba hacia abajo vemos que un padre puede tener N hijos, pero un hijo solo puede tener 1 padre.

El modelo jerárquico no permite relaciones de muchos a muchos o sea de N:M.

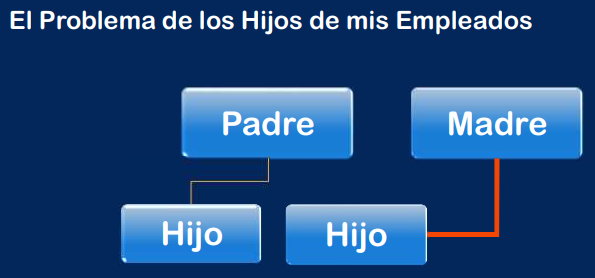
Recapitulando, el modelo jerárquico:

* Es un modelo de árbol invertido
* Tienen 3 niveles (Raíz, Ramas, Hojas )
* Es sumamente simple
* Es muy rígido
* No permite relaciones N:M

El problema que las bases de datos jerárquicas NO pueden solucionar:

Si en mi empresa trabajan el padre y la madre de un hijo no puedo representar eso de forma sencilla.

Tengo que duplicar la información y la redundancia de los datos es la madre de todos los problemas en una base de datos.



De todas formas, este modelo está presente en muchos ejemplos de la vida:

* Directorio de un disco ( Windows Explorer)
* Directorio de iTunes
* Menues de aplicaciones
* Índices de los libros

Estructuras de HTML y JSON

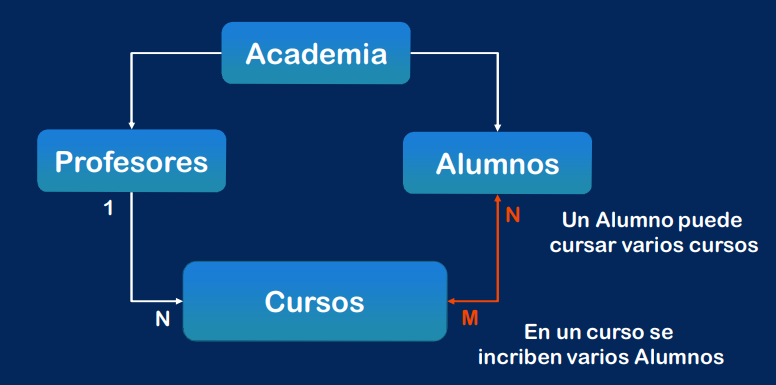


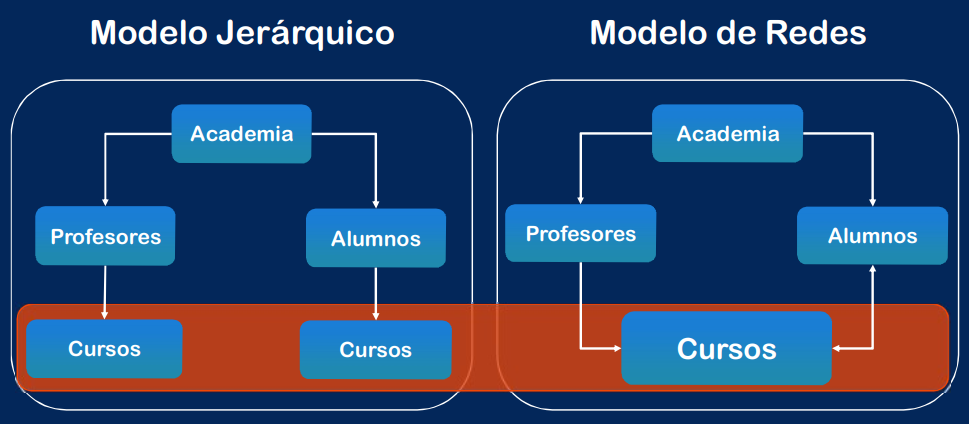
Hay sistemas como IMS y ADABAS que aún usan esta estructura.

3) Modelo de redes

Nace para corregir los defectos del modelo jerárquico y permitir relaciones de N:M.

Se basa en la teoría de los grafos, fue un modelo desafortunado que no se utiliza más.





No necesitamos tener dos entidades “cursos” como en el modelo jerárquico por lo que eliminamos redundancia.

Es mas flexible que el modelo jerárquico y permite representaciones más complejas.

Al eliminar la duplicidad de los datos reduce los costos de almacenamiento e incrementa la velocidad de acceso, por otra parte, ayuda a mantener la integridad de los datos.

Integridad de datos: Mantener y asegurar la consistencia durante toda la vida de los datos.

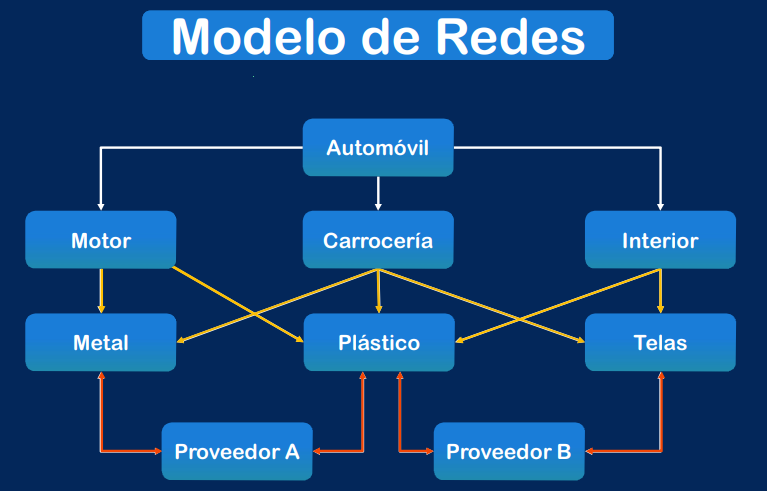
Es preferible no tener datos para tomar una decisión que tener datos errados.

La definición de modelo de redes es MUY COMPLEJA, utiliza apuntadores que son DIFICILES DE MANTENER.

Resumen:

ventajas

Se basa en grafos

Permite modelar relaciones complejas y relaciones N:M

Es un modelo conceptualmente simple

Desventajas:

Es estructuralmente complicado.

Se basa en el manejo de apuntadores que son difíciles de mantener.

Fue presentado en 1969 y en 1970 Codd presenta el modelo relacional el cual logra en 1980 su especificación.

4) Modelo Relacional

E. Codd en el año 1969 escribe un abstracto donde podemos extraer:

“Contendrán muchas Relaciones […]

No será inusual que este conjunto de

Relaciones sea Redundante.”

“Los responsables del control del banco de

Datos deben saber de ellas (redundancias) y

deben tener forma de detectar cualquier

Inconsistencia ‘lógica’ ”

“El Chequeo de Consistencia puede ser útil

para encontrar cambios no autorizados (y

posiblemente fraudulentos) en el contenido

del banco de datos.”

Nos llama la atención sobre 3 items:

* Redundancia
* Inconsistencia lógica
* Chequeo de consistencia

Una visión relacional de los datos:

* Los datos deben describirse en su estructura natural

(Sin imponer estructuras adicionales para representarse en una máquina, el modelo de datos debe ser independiente del dispositivo)

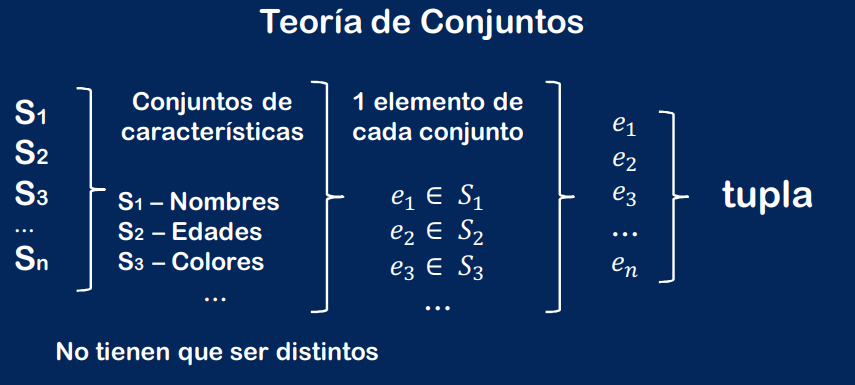
* Provee las bases para un lenguaje de consulta de alto nivel

(Que debe brindar independencia de la estructura de los datos de la máquina, habla de la forma de acceder a los datos)

* Sienta unas bases estables para el tratamiento de la Derivabilidad, Redundancia y Consistencia de las Relaciones.
* Permite evaluar claramente el ámbito, las limitaciones lógicas y las bondades del sistema de información.

Las bases del modelo relacional vienen de la teoría de los conjuntos y del álgebra de conjuntos (las operaciones que se pueden hacer con los conjuntos).

Teoría de conjuntos:



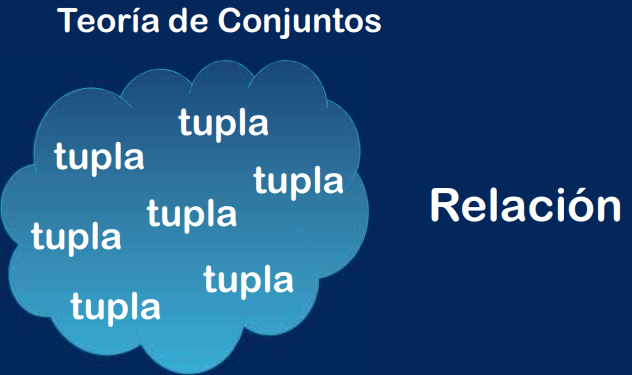
Tenemos un grupo de conjuntos S1 a Sn

* S1 Nombres
* S2 Edades
* S3 Colores

A estos conjuntos de características las llamamos “dominios” y no tienen por que ser diferentes entre si, podemos tener dos conjuntos de nombres o dos conjuntos de colores.

De cada conjunto tomamos un solo elemento y formamos un conjunto nuevo e1, e2 ….en al que vamos a llamar “tupla”

Una tupla es un conjunto de elementos tomados de diferentes dominios (era la fila de una tabla donde cada columna es un dominio).

Formamos un conjunto de tuplas y a este conjunto lo llamamos una relación.

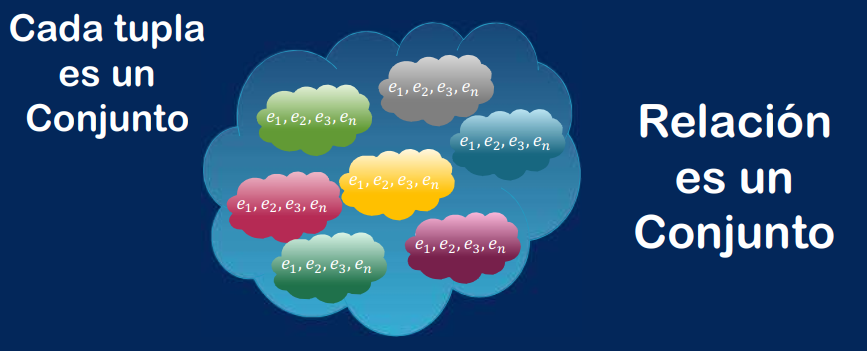
De allí el nombre del modelo

“Modelo Relacional”

Nos hemos acostumbrado a ver las relaciones como una tabla:



No debemos olvidar que una relación es un conjunto de tuplas y una tupla es un conjunto de atributos



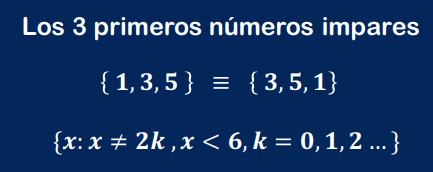
La representación tabular nos da una falsa sensación de orden que no existen en el modelo relacional

Por que es importante recordar que estamos trabajando con conjuntos:

* Un conjunto puede no tener elementos -> Conjunto vacio
  + Una relación puede no tener ninguna tupla y seguir siendo una relación
* Un conjunto queda definido por sus miembros
  + Los números pares
  + Los alumnos de la academia
  + Las personas que están tomando este curso

Esto nos lleva a las siguientes conclusiones:

* Cambiar el orden de un conjunto no define un nuevo conjunto



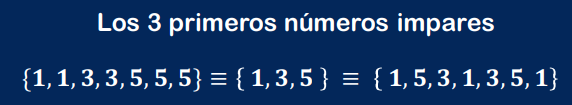
Los conjuntos NO TIENEN ORDEN

Las relaciones NO TIENEN ORDEN simplemente porque contienen las mismas tuplas



Agregando elementos repetidos a un conjunto NO ESTAMOS DEFINIENDO UN NUEVO CONJUNTO

Los elementos repetidos se tratan como un solo elemento



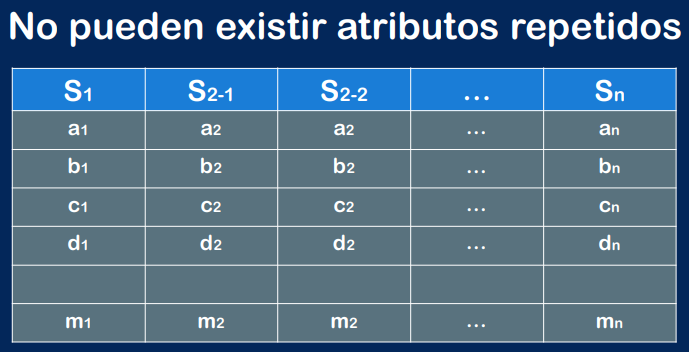
Una característica muy importante del modelo relacional es que

* ¡NO PUEDE HABER DOS TUPLAS REPETIDAS!
* ¡NO PUEDE HABER DOS ATRIBUTOS REPETIDOS!





Dos atributos pueden tener el mismo DOMINIO pero las columnas deben llamarse distinto:



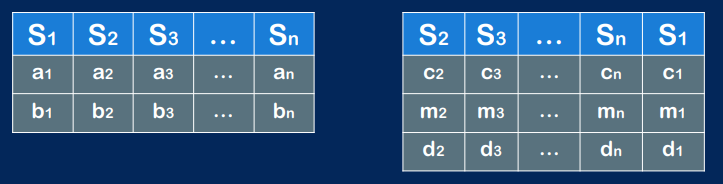
Definimos algunos términos de la teoría de conjuntos:

**Inclusión:** Todos los elementos de un Conjunto “A” son parte de los elementos de otro Conjunto “B” A está incluido en B. A es sub conjunto de B

Podemos tener relaciones que sean sub conjunto de otras relaciones

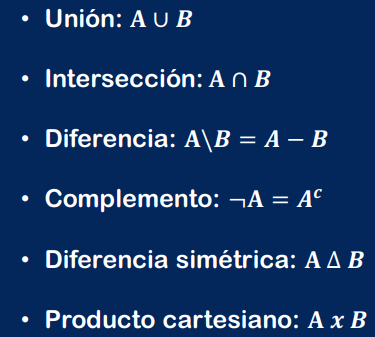


**Disyunción**: Cuando dos conjuntos no tienen ningún elemento en común. Dos relaciones son disyuntas cuando no tienen ninguna tupla en común.



**Operaciones**: Si se aplica una operación sobre un conjunto el resultado siempre es otro conjunto

Cualquier operación que apliquemos sobre una relación el resultado es otra relación



Lógica de Predicados: Es otra de las bases del modelo relacional y según Cobb se basa en la lógica de segundo orden. Provee las bases para un lenguaje de consulta de alto nivel.

Un predicado es una expresión (o función) que dependiendo de sus valores puede ser evaluada a un resultado “verdadero” o “falso”

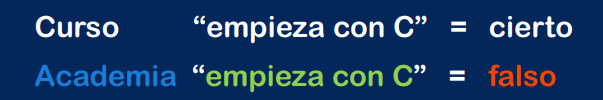
Es una expresión que se evalúa a un valor de verdad.

Ejemplo: Expresión que se evalúa hacia un valor de verdad

• Palabras que empiezan con la letra C

• Alumnos inscritos en el curso de Diseño

• Clientes que cumplen años en Febrero



Evaluamos los 3 elementos de un predicado:



