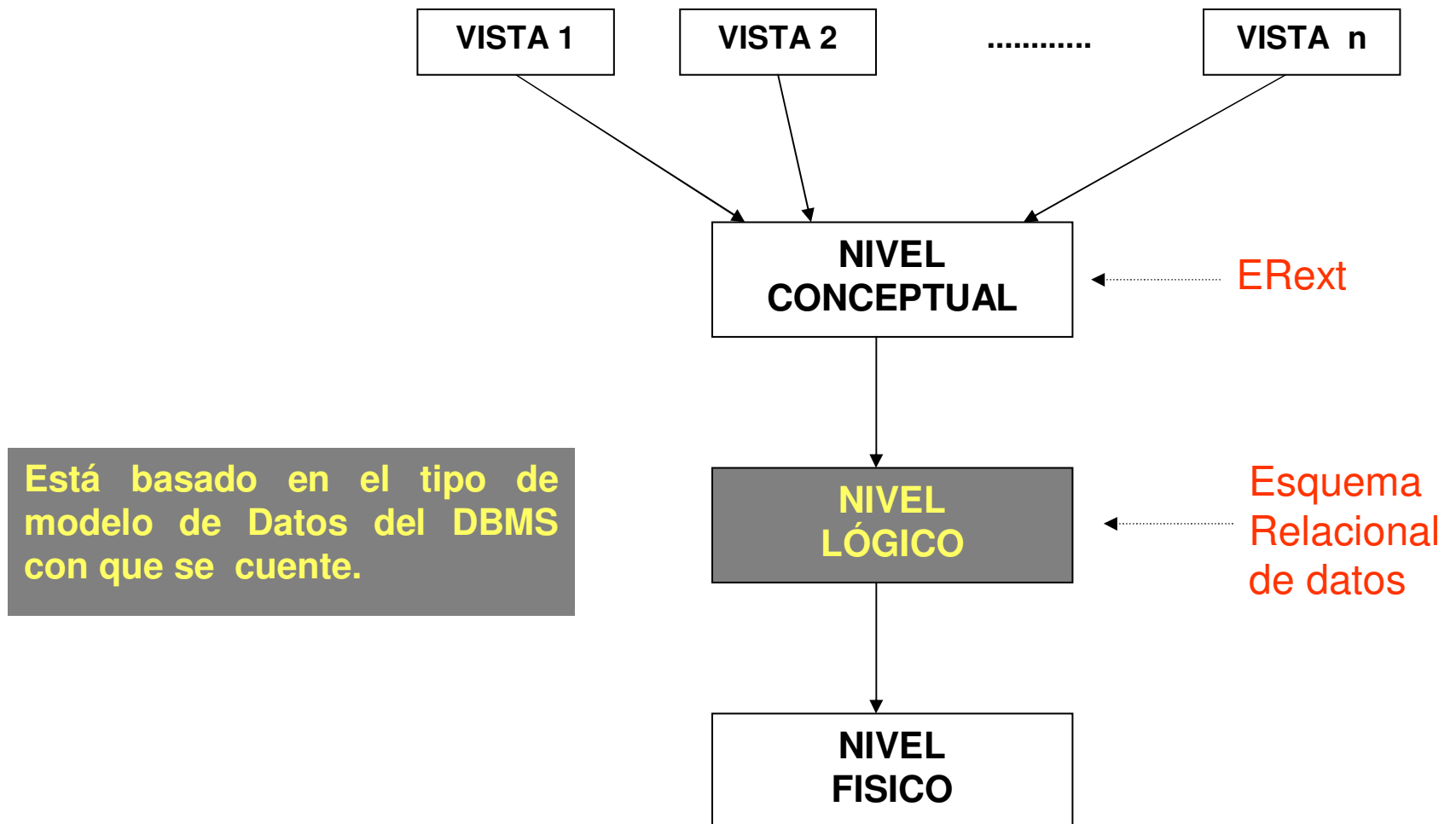


Diseño Lógico: Modelo Relacional

Previo:

Abstracción : Consideración de lo esencial, omisión de las diferencias.

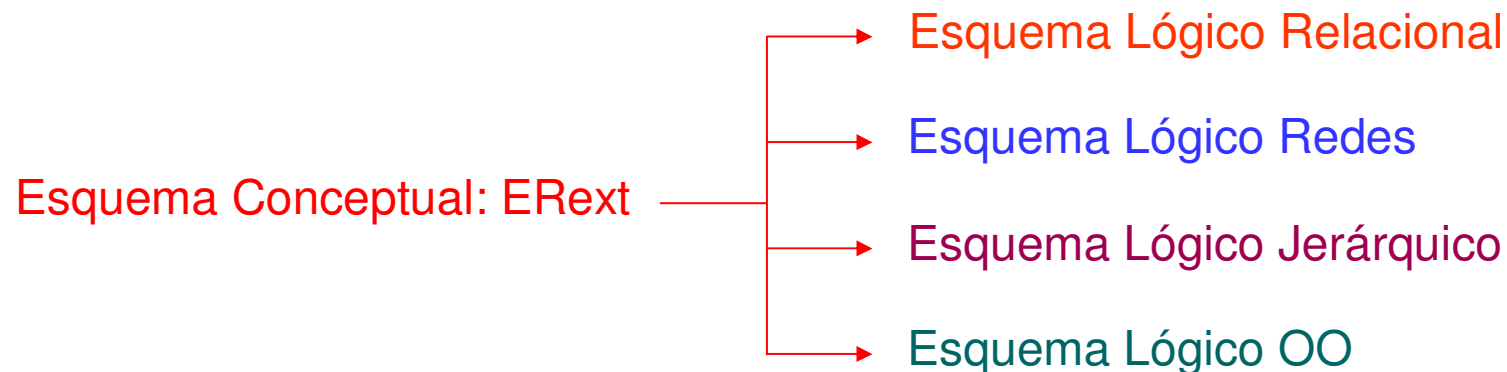


Diseño Lógico

El diseño lógico es el proceso de construir un esquema que representa el dominio de conocimiento y las reglas del negocio, basándose en un modelo de base de datos específico.

En esta etapa, se transforma el esquema conceptual en un esquema lógico que utilizará las estructuras de datos del modelo de base de datos en el que se basa el SGBD que se vaya a utilizar. Conforme se va desarrollando el esquema lógico, éste se va probando y validando con los requisitos de usuario.

Un esquema lógico es una descripción de la estructura de la base de datos en términos de las estructuras de datos que puede procesar un tipo de SGBD. Un modelo lógico es un formalismo usado para especificar esquemas lógicos (modelo relacional, modelo de red, OO, etc.).



MODELO DE DATOS

Un modelo de datos es un conjunto de conceptos que sirven para describir la estructura de una base de datos: los datos, las relaciones entre los datos y las restricciones que deben cumplirse sobre los datos. Los modelos de datos contienen también un conjunto de operaciones básicas para la realización de consultas (lecturas) y actualizaciones de datos. Además, los modelos de datos más sofisticados incluyen conceptos para especificar comportamiento, permitiendo especificar un conjunto de operaciones definidas por el usuario

Un modelo se expresa a través de:

- a.- **La estructura.** Un modelo de datos define las reglas bajo las cuales los datos están estructurados. Sin embargo este no provee una interpretación completa acerca del significado de los datos y la manera como deberán ser usados.
- b.- **Las restricciones** asociadas con los datos. Ellas son “aseveraciones” (siempre son verdad) para cualquier estructura dentro del modelo de datos para las cuales se aplica. Estas pueden ser de dos tipos: implícito o explícito.
- c.- **Las operaciones** permitidas sobre los datos; estas operaciones están normalmente relacionadas con la estructura de los datos, por ejemplo las operaciones son ejecutadas dentro del contexto provisto por las estructuras (un modelo de datos define reglas generales para la especificación de las estructuras de los datos y las operaciones permitidas sobre los mismos).

La Estructura en el Modelo Relacional

El elemento básico del modelo es la *relación* y un modelo de base de datos relacional es una colección de relaciones.

MODELO RELACIONAL

El Modelo Relacional de Edward Codd, se presentó en un artículo de 1970 llamado "A Relational Model of Data for Large Shared Data Banks" (Communication of the ACM, vol. 13, nº 6 junio 1970). Está considerado una de las contribuciones más importantes dentro del desarrollo de la tecnología de bases de datos. Dentro de las cualidades del modelo, que lo han hecho el modelo más utilizado en los SGBD comercializados, están:

- Simplicidad de los conceptos.
- Operadores de manipulación de la data muy poderosos.
- Apoyo de una sólida base matemática.

La Estructura en el Modelo Relacional

ESTUDIANTE		ASIGNATURA		APROBO	
c.i.	nombre	código	nombre-asig.	c.i.	código-mat
11	Marlene Dietrich	C01	Matemáticas I	11	C01
22	María Félix	C02	Algorit. Prog. I	33	C02
33	Pedro Almodovar	C03	DHP	55	C01
44	Ingmar Berman	C04	Matemát. Disc.		
55	Doris Wells				

Atributo: columna de la tabla.

Dominio de un atributo: todos los posibles valores que puede tomar ese atributo.

Tupla: fila de la tabla.

Grado de una relación: es el número de atributos que contiene.

La cardinalidad de una relación: es el número de tuplas que contiene.

Una base de datos relacional es un conjunto de **relaciones normalizadas**.

La Estructura en el Modelo Relacional

Si consideramos:

ASIGNATURA con **codigo** : {C01,C02,C03,C04,...}

nombre_asig : {Matemáticas I, Algorit. Prog I, DHP, Matemát. Disc.,...}

Cada código_i es una instancia en particular de **codigo** y cada **nombre_asig**_j es una instancia en particular de **nombre_asig**.

Entonces una **tupla** (C01,Matemáticas I) ∈ conjunto de m tuplas de **ASIGNATURA** donde m es llamada la cardinalidad de **ASIGNATURA**. Se puede decir que:

$$ASIGNATURA \subseteq \prod_{i=1}^n D_i, \quad n = 2, \quad n \text{ es llamado grado de la relación.}$$

donde $D_1 = \text{Código}$ y $D_2 = \text{Nombre_asig}$.

Definición matemática de Relación

"Dada una colección de conjuntos D_1, D_2, \dots, D_n (no necesariamente distintos), R es una *relación* sobre esos n conjuntos si R es un conjunto ordenado de m tuplas $\langle d_1, d_2, d_3, \dots, d_n \rangle$ donde $d_1 \in D_1, d_2 \in D_2, \dots, d_n \in D_n$. Los conjuntos D_1, D_2, \dots, D_n se llaman dominios de R y el valor de n es el grado de R ".

La única diferencia es que, en el contexto de la bases de datos, se deben colocar nombres a los atributos y, no necesariamente, las tuplas que pertenecen a una relación se encontrarán ordenadas. Otro elemento, es que la tupla producto de $D_1 \times D_2$ es igual a la tupla producto de $D_2 \times D_1$, no siendo así, en el ámbito de las matemáticas.

La *relación* es entonces un conjunto de elementos llamados tuplas y puede definirse por intensión y por extensión.

Definición por Intensión

Se da el predicado P de una tupla de R tal que:

$$R = \{[a_1, a_2, \dots, a_n] / P(a_1, a_2, \dots, a_n) \text{ es verdadero}\}$$

APROBO	
c.i.	código-mat
11	C01
33	C02
55	C01

La relación: **APROBO(c.i,codigo_mat)** se puede definir como el predicado por intensión de la siguiente forma:

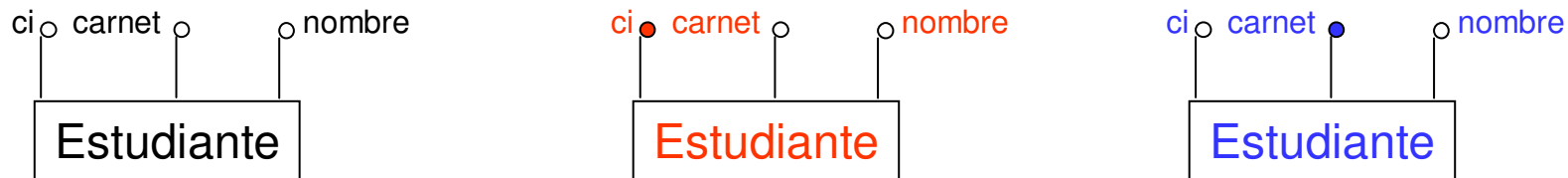
P(cí,código-mat): "el estudiante *cí* aprobó la asignatura *codigo-mat*".

Y se puede definir por extensión: "la estudiante 11 aprobó la asignatura C01"

La representación tabular es una definición por extensión de la relación.

Las Claves en el Modelo Relacional

Al conjunto mínimo de atributos que puede identificar a cada tupla dentro de la extensión de una relación, se le denomina *clave candidata*.



Pueden haber varias claves candidatas en una relación. Una de ellas debe ser escogida para ser la *clave primaria*. Al resto de las claves candidatas se les llama *claves alternas* de la relación. Por ejemplo:

ESTUDIANTE(ci,carnet,nombre)

Clave alterna: carnet

Por otra parte, se dice que un *dominio es primario* si sobre él se ha definido una clave primaria.

Las Claves en el Modelo Relacional

Una *clave foránea*, en una relación R, es un atributo definido sobre un dominio primario que no es el dominio de la clave primaria de R (sino de otra relación).

Por ejemplo:

ESTUDIANTE(ci,carnet,nombre)
ASIGNATURA(codigo_mat,nombre_asig)
APROBO(ci,codigo_mat)

- 1) *ci* es *clave foránea* en **APROBO** ya que su dominio es un *dominio primario* en **ESTUDIANTE**.
- 2) *codigo_mat* es *clave foránea* en **APROBO** ya que su dominio es un *dominio primario* en **ASIGNATURA**.

Características Generales de las Relaciones

- Cada relación tiene un nombre y éste es distinto del nombre de todas las demás.
- No hay dos atributos que se llamen igual.
- El orden de los atributos no importa: los atributos no están ordenados.
- Cada tupla es distinta de las demás: no hay tuplas duplicadas.
- El orden de las tuplas no importa: las tuplas no están ordenadas.

Tipos de Relaciones

- **Relaciones base.** Son relaciones reales que tienen nombre y forman parte directa de la base de datos almacenada (son autónomas).
- **Vistas.** También denominadas relaciones virtuales, son relaciones con nombre y derivadas: se representan mediante su definición en términos de otras relaciones con nombre, no poseen datos almacenados propios.
- **Instantáneas.** Son relaciones con nombre y derivadas. Pero a diferencia de las vistas, son reales, no virtuales: están representadas no sólo por su definición en términos de otras relaciones con nombre, sino también por sus propios datos almacenados. Son relaciones de sólo de lectura y se refrescan periódicamente.

Tipos de Relaciones

- **Resultados de consultas.** Son las relaciones resultantes de alguna consulta especificada. Pueden o no tener nombre y no persisten en la base de datos.
- **Resultados intermedios.** Son las relaciones que contienen los resultados de las subconsultas. Normalmente no tienen nombre y tampoco persisten en la base de datos.
- **Resultados temporales.** Son relaciones con nombre, similares a las relaciones base o a las instantáneas, pero la diferencia es que se destruyen automáticamente en algún momento apropiado.

Restricciones en el Modelo Relacional

Las restricciones que tiene el modelo relacional se denominan restricciones de integridad. Estas restricciones deben ser especificadas, en forma explícita, en el esquema relacional. Se espera que estas restricciones se cumplan para cada caso (u ocurrencia) del esquema de la base de datos, de forma de garantizar la integridad de la data.

La integridad se divide en:

- a) **Integridad del dominio:** se refiere al tipo de definición del dominio. Controla la sintaxis y la semántica de cualquier dato. Puede ocurrir que sobre un dominio se definen atributos que, semánticamente, sean diferentes pero que sean, a su vez, equivalentes dado que poseen el mismo dominio.

Por ejemplo, considere la relación:

VIAJE(num_viaje, num_zona, fecha_sal, fecha_lleg, ciudad_destino)

Restricciones en el Modelo Relacional

Podrían definirse los siguientes dominios:

Número: entero de cinco dígitos.

Fecha: fecha definida con un formato 'date' (dd/mm/aa).

Ciudad: *string* de 20 caracteres.

Entonces, los ítems que conforman los atributos de la relación **Viaje**, se definen de la siguiente manera:

num_viaje se define sobre Número	←	el mismo dominio
num_zona se define sobre Número	←	
fecha_sal se define sobre Fecha	←	el mismo dominio
fecha_lleg se define sobre Fecha	←	
ciudad_destino se define sobre Ciudad		

b) Integridad de la clave y de la relación: se refiere a definir las claves primarias de cada relación y cómo deben ser los valores de los atributos que conforman la clave primaria de una relación.

Las restricciones de la clave especifican cuáles atributos son la clave primaria en una relación, si observamos la figura, podemos decir que (ci) es la clave de ESTUDIANTE, (código) es la clave de ASIGNATURA y (ci,código_mat) es la clave de APROBO.

ESTUDIANTE		ASIGNATURA		APROBO	
c.i.	nombre	código	nombre-asig.	c.i.	código-mat
11	Marlene Dietrich	C01	Matemáticas I	11	C01
22	María Félix	C02	Algorit. Prog. I	33	C02
33	Pedro Almodovar	C03	DHP	55	C01
44	Ingmar Berman	C04	Matemát. Disc.		
55	Doris Wells				

Por otro lado, las restricciones sobre la relación dicen que cada relación posee exactamente una sola clave primaria y que, ésta, no tiene dos valores iguales para referirse a dos tuplas diferentes. Los valores deben ser únicos y no nulos (por nulo se entiende cualquier carácter que represente al conjunto vacío).

Si una clave pudiera tener un valor nulo, significaría que alguna tupla no puede identificarse.

c) **Integridad Referencial**: se refiere a los valores que puedan tomar los atributos de la clave primaria de una relación y los valores que estos mismos atributos puedan tomar, como clave foránea, en otra relación. Por ejemplo, en la figura, la existencia de las relaciones **ESTUDIANTE** y **ASIGNATURA** no depende de ninguna otra relación (por eso, a estas relaciones se les llaman primarias).

ESTUDIANTE		ASIGNATURA		APROBO	
c.i.	nombre	código	nombre-asig.	c.i.	código-mat
11	Marlene Dietrich	C01	Matemáticas I	11	C01
22	María Félix	C02	Algorit. Prog. I	33	C02
33	Pedro Almodovar	C03	DHP	55	C01
44	Ingmar Berman	C04	Matemát. Disc.		
55	Doris Wells				

En **ESTUDIANTE** ci, es la clave primaria y en **ASIGNATURA** código es la clave primaria. Sin embargo, la existencia de la relación **APROBO** *depende de la existencia* de **ESTUDIANTE** y de **ASIGNATURA** (a este tipo de relación se les llama secundarias). Debido a esto, el conjunto de valores que toma el atributo ci, en **APROBO** debe ser un subconjunto del conjunto de valores de ci en **ESTUDIANTE**, de igual forma sucede con el atributo código-mat de **APROBO** y código de **ASIGNATURA**.

Operaciones en el Modelo Relacional

Las operaciones que se efectúan sobre las relaciones son: creación, inserción y actualización.

Estas se escriben en el Lenguaje de Manipulación de Datos (LMD).

Los LMD se dividen en formales y comerciales y entre los más usados se encuentran:

- Álgebra Relacional.
- Cálculo Relacional.
- Structured Query Language (SQL).
- Query By Example (QBE).