

El Modelo Relacional. Dinámica

TEMA IV

© Grupo de Bases de Datos Avanzadas – Univ. Carlos III de Madrid

Modelo Relacional.
Dinámica

Índice

DINÁMICA DEL MODELO RELACIONAL

- IV.1 INTRODUCCIÓN
- IV.2 ALGEBRA RELACIONAL
- IV.3 TRATAMIENTO DE VALORES NULOS
- IV.4 CÁLCULO RELACIONAL
- IV.5 OPTIMIZACIÓN DE CONSULTAS

Tema IV

FBD3 IV.2

© Grupo de Bases de Datos Avanzadas – Univ. Carlos III de Madrid

- ❑ La componente dinámica del Modelo Relacional permite la transformación entre estados de la BD.
- ❑ Se entiende una *relación* como un conjunto (de tuplas sin repetición)
- ❑ Una BD es una colección de esos conjuntos (más alguna restricción). Si las operaciones están alterando las relaciones en el tiempo, se puede decir que son cambios en el estado de la Base de Datos (transiciones):

$$O(BD_i) = BD_j \quad * \text{ la base de datos cambia de estado}$$

- ❑ Para obrar esos cambios, es necesario intervenir a través de un lenguaje.

- ❑ Los lenguajes relacionales son de especificación (frente a los navegacionales). Se distinguen dos tipos:
 - Algebraicos: los cambios se describen mediante las operaciones que sufre la BD → **álgebra relacional**
 - Predicativos: se describe el estado final de la BD deseado, sin especificar cómo alcanzarlo → **cálculo relacional**, se dividen a su vez en dos tipos:
 - Orientados a tuplas
 - Orientados a dominios

- ❑ El **Álgebra Relacional** lo componen un conjunto de operaciones extraídas del **Álgebra de Conjuntos**. Estos operadores (O) se aplicarán sobre relaciones (conjuntos) y darán lugar a nuevas relaciones (nuevos conjuntos):

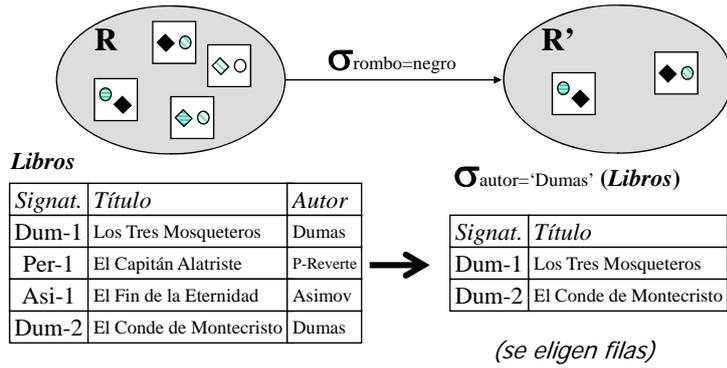
$$O: \text{esq}R_1 \times \text{esq}R_2 \rightarrow \text{esq}R_3 \quad ,, O(R_1, R_2) = R_3$$

- ❑ Los operadores se aplican a una extensión r de una relación R, no a un esquema de relación.
- ❑ Los operadores definidos permitirán expresar la componente de *selección* de las operaciones del LMD.

CLASIFICACIONES DE OPERADORES

- ❑ CODD distinguió dos grupos con cuatro operadores cada uno:
 - Operadores **tradicionales** de conjuntos: **unión**, **intersección**, **diferencia** y **producto cartesiano**.
 - Operadores **relacionales** especiales: **selección**, **proyección**, **combinación** y **división**.
- ❑ También pueden dividirse en:
 - **Operadores primitivos**: aquellos operadores **esenciales** que no pueden obtenerse a partir de otros (convierten al álgebra relacional en un **lenguaje completo**)
 - **Operadores derivados**: pueden obtenerse **aplicando** varios de los **primitivos**.
- ❑ Por último, pueden clasificarse según en número de operandos:
 - **Unarios**: el operador tiene una **única relación** como operando
 - **Binarios**: el operador tiene **dos relaciones** como operandos.

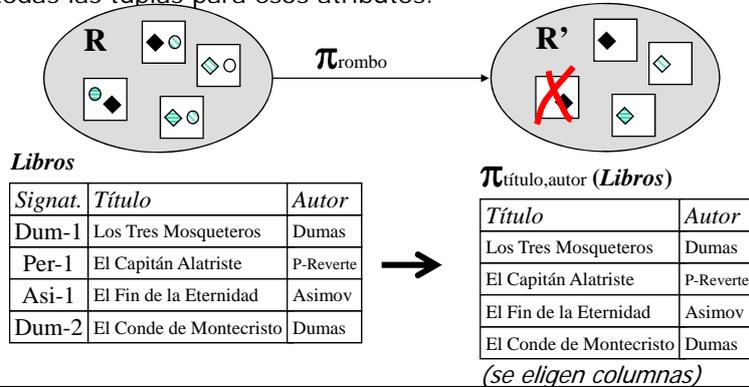
SELECCIÓN (σ): Obtención del subconjunto de tuplas de una relación que satisface una determinada expresión lógica o *predicado de selección*.



Tema IV

FBD3 IV.7

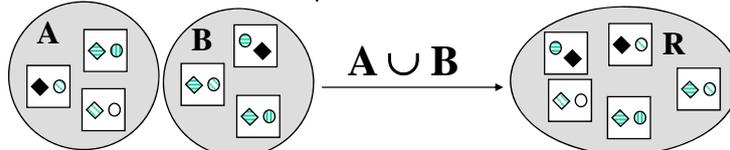
PROYECCIÓN (π): Se define sobre un subconjunto de atributos de una relación, seleccionando únicamente valores no repetidos en todas las tuplas para esos atributos.



Tema IV

FBD3 IV.8

UNIÓN(\cup): Conjunto de todas las tuplas (sin repetición) incluidos en ambas relaciones. El esquema de relación no se modifica.



Aventuras

Signat.	Título	Autor
Dum-1	Los Tres Mosqueteros	Dumas
Per-1	El Capitán Alatraste	P-Reverte
Dum-2	El Conde de Montecristo	Dumas

Lecturas

Signat.	Título	Autor
Ner-1	Veinte poemas de amor	Neruda
Asi-1	El Fin de la Eternidad	Asimov

Aventuras \cup Lecturas

Signat.	Título	Autor
Dum-1	Los Tres Mosqueteros	Dumas
Per-1	El Capitán Alatraste	P-Reverte
Asi-1	El Fin de la Eternidad	Asimov
Dum-2	El Conde de Montecristo	Dumas
Ner-1	Veinte poemas de amor	Neruda

DIFERENCIA($-$): Conjunto de todas las tuplas que pertenecen a una relación pero no a la otra.



Aventuras

Signat.	Título	Autor
Dum-1	Los Tres Mosqueteros	Dumas
Per-1	El Capitán Alatraste	P-Reverte
Dum-2	El Conde de Montecristo	Dumas

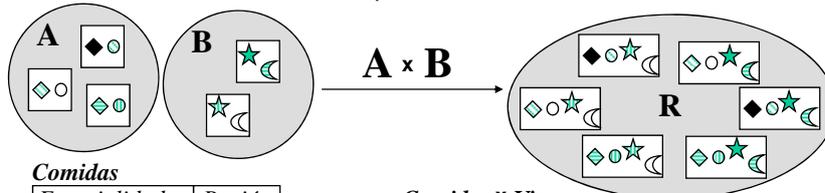
Lecturas

Signat.	Título	Autor
Ner-1	Veinte poemas de amor	Neruda
Per-1	El Capitán Alatraste	P-Reverte

Aventuras - Lecturas

Signat.	Título	Autor
Dum-1	Los Tres Mosqueteros	Dumas
Dum-2	El Conde de Montecristo	Dumas

PRODUCTO CARTESIANO (X): Conjunto formado por la concatenación de todas las tuplas de ambas relaciones.



Comidas

Especialidad	Región
Gazpacho	Andalucía
Cocido	Castilla
Pote	Galicia

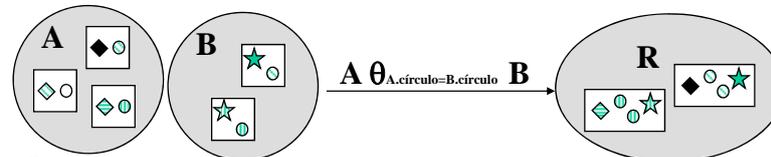
Vinos

Denominación
Rioja
Rueda

Comidas x Vinos

Especialidad	Región	Denominación
Gazpacho	Andalucía	Rioja
Gazpacho	Andalucía	Rueda
Cocido	Castilla	Rioja
Cocido	Castilla	Rueda
Pote	Galicia	Rioja
Pote	Galicia	Rueda

COMBINACIÓN (θ): Tuplas del producto cartesiano que cumplen una condición determinada.



Personas

Nombre	Edad
Fulano	29
Mengano	18
Zutano	9

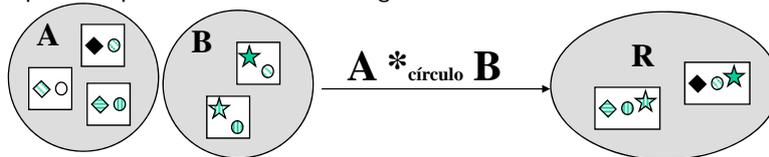
Películas

Título	Calificación
Matrix	18
Sólo tu	0

Personas θ_{Edad ≥ Calificación} Películas

Nombre	Edad	Título	Calificación
Fulano	29	Matrix	18
Fulano	29	Sólo tu	0
Mengano	18	Matrix	18
Mengano	18	Sólo tu	0
Zutano	9	Sólo tu	0

COMBINACIÓN NATURAL ($*$): Tuplas del producto cartesiano que cumplen la condición de igualdad.



Personas

Nombre	DNI
Fulano	569074
Mengano	434848
Zutano	383090

Coches

Marca	Dueño
Ford Escort	383090
Seat Ibiza	569074

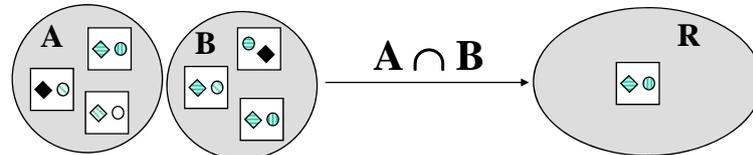


Personas * DNI Coches

Nombre	DNI	Marca
Fulano	569074	Seat Ibiza
Zutano	383090	Ford Escort

• observar que se elimina la columna repetida

INTERSECCIÓN (\cap): Conjunto de tuplas que pertenecen a ambas relaciones.



Aventuras

Signat.	Título	Autor
Dum-1	Los Tres Mosqueteros	Dumas
Per-1	El Capitán Alatraste	P-Reverte
Dum-2	El Conde de Montecristo	Dumas

Lecturas

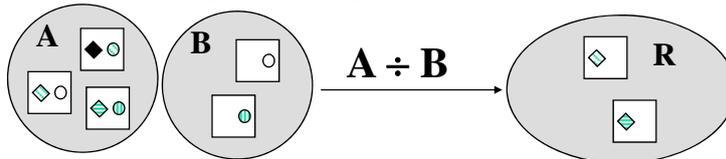
Signat.	Título	Autor
Ner-1	Veinte poemas de amor	Neruda
Per-1	El Capitán Alatraste	P-Reverte



Aventuras \cap Lecturas

Signat.	Título	Autor
Per-1	El Capitán Alatraste	P-Reverte

DIVISIÓN(⋔): Conjunto de tuplas de la primera (R1) que contienen a la segunda (R2) y sin ésta última.



Libros

Signat.	Título	Autor
Dum-1	Los Tres Mosqueteros	Dumas
Per-1	El Capitán Alatriste	P-Reverte
Dum-2	El Conde de Montecristo	Dumas

Autores

Autor
Dumas

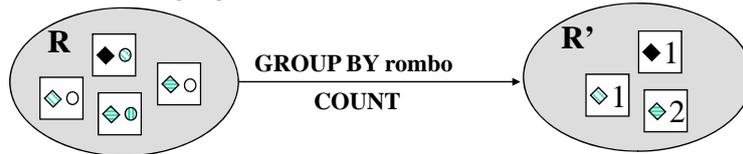
Libros ÷ Autores

Signat.	Título
Dum-1	Los Tres Mosqueteros
Dum-2	El Conde de Montecristo

OPERADORES DERIVADOS EXPRESADOS A TRAVÉS DE OPERADORES PRIMARIOS

Operador Derivado	Expresión a través de Primarios
Intersección	$A - (A - B)$
División	$\pi \dots - \dots (A' \times B')$
Combinación	$\sigma \dots (A \times B)$
Combinación Natural	$\pi \dots \sigma \dots (A \times B)$

AGRUPACIÓN ('Group By'): Permite reunir tuplas con valores comunes en subconjuntos, de manera que puedan aplicarse funciones de agregación.



Personas

Nombre	Edad	Nacionalidad
Fulano	29	Español
Mengano	49	Español
John Doe	73	Inglés
Smith	14	Inglés
Zutano	3	Español
Pelancejo	25	Español

Group by Nacionalidad (*Personas*)

Nacionalidad	...	Avg(Edad)	Count	...
Español	...	26,5	4	...
Inglés	...	43,5	2	...

CIERRE TRANSITIVO: Consiste en agregar a la relación todas las tuplas que se pueden derivar por transitividad.



Personas (Nombre, Edad, Hijo)

Nombre	Edad	Hijo
Fulano	29	Zutano
Mengano	49	Pelancejo
Zutano	3	
Pelancejo	25	Zutano

Cierre Hijo (*Personas*)

Nombre	Edad	Hijo
Fulano	29	Zutano
Mengano	49	Pelancejo
Mengano	49	Zutano
Zutano	3	
Pelancejo	25	Zutano

VALOR NULO

- Señal** o marca utilizada para representar información desconocida, inaplicable, inexistente, no válida, no proporcionada, etc.
- Ejemplo:
coche(matrícula, marca, modelo, ..., tipo de carburador, ...)
 - puede que no sepa de qué tipo es el carburador, pero que me interese mantener el resto de la información
 - algunas ocurrencias pueden no tener carburador (ni tipo), pero deben tener cabida en la relación
 - si quiero añadir un atributo nuevo a la relación, ¿qué valor deben adoptar las tuplas existentes para ese nuevo atributo?
- Para cubrir esta necesidad se considera un símbolo especial: el *valor nulo*. Pero ¿cómo se opera con él?

- comparación** (de un valor con un valor nulo): el resultado es *impredecible*.
Esto afecta a las operaciones lógicas: ya no hay sólo dos valores (T/F), sino que se añade el 'impredecible' (Q)
Se introduce la **lógica trivaluada**, con los siguientes operadores:

AND	T	Q	F
T	T	Q	F
Q	Q	Q	F
F	F	F	F

OR	T	Q	F
T	T	T	T
Q	Q	T	Q
F	F	T	Q

NOT	T	F
T	F	
Q	Q	
F	T	

- El problema de la lógica trivaluada es su comportamiento con tautologías
Ejemplo: (edad<30) OR (edad>30) OR (edad=30) ≠ 1 !!
- Para poder precisar el valor indecible, se introduce la **lógica cuatrivaluada**, que desdobra ese valor en otros dos: *desconocido* e *inaplicable*.

- ❑ **operaciones algebraicas:** hay que decidir cómo tratar lo impredecible (cómo seleccionar condiciones impredecibles, cuándo dos tuplas son la misma, etc.)
- ❑ **operaciones aritméticas** (+, -, *, ÷):
el resultado de operar cualquier cosa con un valor nulo es **nulo**
- ❑ **operaciones de agregación:** hay que decidir cómo agregar lo nulo (por defecto, se cuenta como una tupla que no afecta en la operación de agregación, salvo para la operación de calcular la cardinalidad)
 - Count: cardinalidad de la relación (contando nulos como un valor)
 - Máx (Min): el máximo del resto (sin nulos) o nulo.
 - Avg: la media del resto (sin nulos) o nulo
 - ...
- ❑ **operaciones especiales :**
 - **Is Null:** comprueba que el valor cuestionado sea nulo
 - **If Null:** derivada de la anterior, señala un valor para sustituir nulos.

- ❑ Propuesto por CODD como alternativa al álgebra.
- ❑ Es un lenguaje **predicativo**: sólo es necesario indicar el resultado a obtener, expresado mediante predicados de primer orden.
- ❑ Pueden ser de dos tipos:
 - **Orientados a tuplas:** se definen variables que hacen referencia a tuplas de la relación
 - **Orientados a dominios:** se definen variables que representan valores de un dominio

Lenguajes orientados a tuplas

- Las variables se asocian a tuplas de la relación y abarcan una determinada relación de la base de datos.
- Las constantes son valores de los dominios sobre los que se definen los atributos
- Una consulta simple en cálculo de tuplas tiene la forma:

$\{t \mid \text{COND}(t)\}$

Ej.: $\{t \mid \text{EMPLEADO}(t) \text{ and } \text{SALARIO} > 50000\}$

Lenguajes orientados a tuplas

- Una condición o fórmula está formada por átomos conectados mediante los operadores and, or y not.
- Un átomo puede ser:
 - $R(t_i)$
 - $t_i.A \text{ oper } t_j.B$
 - $t_i.A \text{ oper } c \text{ ó } c \text{ oper } t_j.B$
- Los operadores permitidos son: $>, <, >=, <=, =$

Lenguajes orientados a tuplas

- Operadores especiales:
 - Cuantificador Existencial, \exists
 - Cuantificador Universal, \forall
- Si F es una fórmula, $(\exists t)(F)$ también lo es, y toma valor TRUE si F es TRUE para alguna tupla asignada a t.
- Si F es una fórmula, $(\forall t)(F)$ también lo es, y toma valor TRUE si F es TRUE para toda tupla asignada a t.

Lenguajes orientados a tuplas

- Ejemplo de consulta:

Obtener el nombre y la dirección de todos los empleados que trabajan para el departamento de 'Investigación'

$\{t.NombreP, t.Apellido, t.Dirección \mid EMPLEADO(t) \text{ and } ((\exists d) (DEPARTAMENTO(d) \text{ and } d.NombreD='Investigación' \text{ and } d.Número=t.ND))\}$

Lenguajes orientados a tuplas

- ❑ La estructura de las consultas es:
[Definición de variables]
OPERADOR OBJETIVO PREDICADO
- **OPERADOR:** Acción a realizar sobre los datos seleccionados (recuperación, inserción, borrado o modificación)
- **OBJETIVO:** Indica qué atributos de qué relaciones se van a recuperar.
- **PREDICADO:** Condición que deben verificar las tuplas seleccionadas (criterio de selección), construida a partir de los operadores del cálculo relacional.

Lenguajes orientados a tuplas

EJEMPLO

- ❑ Lenguaje ALPHA (no dispone de implementación pero constituyó la base del lenguaje QUEL de INGRES).
- ❑ Los operadores definidos son: GET, PUT, DELETE, DROP, UPDATE, HOLD y RELEASE.
- ❑ Ejemplo de consulta en ALPHA:

GET RESULTADO (AUTOR.nombre, LIBRO.editorial):
AUTOR.nombre = LIBRO.autor

Lenguajes orientados a dominios

- Las variables abarcan valores individuales de los dominios de los atributos
- Si se quiere formar una relación de grado n como resultado de una consulta, serán necesarias n variables de dominio.
- Una consulta simple en cálculo de dominios tiene la forma:

$$\{x_1, x_2, \dots, x_n | \text{COND}(x_1, x_2, \dots, x_n, x_{n+1}, \dots, x_{n+m})\}$$

Lenguajes orientados a dominios

- Los átomos se definen de forma equivalente a los átomos del cálculo relacional de tuplas.
- Se aplican los mismos operadores para combinar los átomos: $>$, $<$, $>=$, $<=$, $=$
- También se consideran los cuantificadores existencial (\exists) y universal (\forall)

Lenguajes orientados a dominios

- ❑ Ejemplo de consulta:
Obtener la fecha de nacimiento y la dirección del empleado cuyo nombre es 'José Pérez'
 $\{uv \mid (\exists q)(\forall r) (\text{EMPLEADO}(qrstuvwxyz) \text{ and } q='José' \text{ and } r='Pérez')\}$
- ❑ El único lenguaje comercial basado en el cálculo relacional de dominios es QBE (Query By Example)

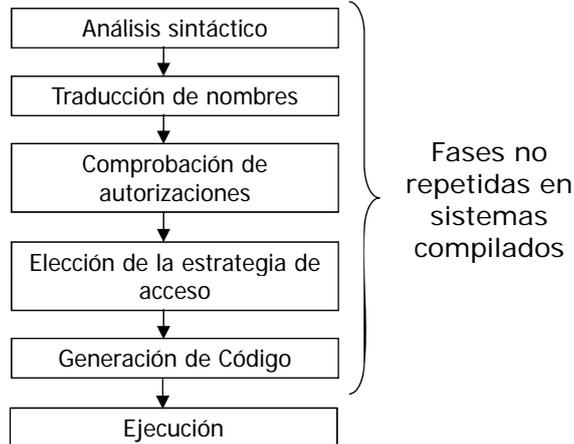
- ❑ Los lenguajes relacionales son poco procedimentales, el usuario indica qué quiere obtener pero no como. Esto implica que el sistema debe escoger el camino más apropiado para acceder a los datos de modo eficiente.
- ❑ **Optimizador:** Módulo incluido en el SGBD capaz de seleccionar una estrategia óptima de acceso a los datos.
- ❑ Es necesario indicar qué parámetros se desean optimizar (compromiso entre recursos utilizados y tiempo de respuesta)
- ❑ Los sistemas relacionales son más fácilmente optimizables que los no relacionales.

- Los sistemas actuales tienden a minimizar el tiempo de respuesta a la consulta, aunque hay que tener en cuenta el ahorro de otros recursos:
 - Costes de comunicación
 - Acceso a almacenamiento secundario
 - Almacenamiento
 - Computación
 - ...
- Todos estos factores intervienen en la función de coste a optimizar.
- Actualmente, los algoritmos de optimización son fijos, el administrador no puede actuar sobre ellos.

Sistemas compilados vs. Sistemas interpretados

- Los sistemas compilados generan código máquina antes de ejecutar la consulta, aumentando el rendimiento.
- En los sistemas interpretados, las consultas se interpretan antes de su ejecución.
- Los productos actuales suelen ofrecer ambas alternativas. Por ejemplo, procedimientos almacenados (Stored Procedures), que se conservan como parte del diccionario de datos.
- En sistemas compilados existe un módulo compilador del LMD, del que forma parte el optimizador de consultas.

Sistemas compilados. Fases de procesamiento de una consulta.



Optimización de consultas simples

- ❑ El procedimiento de optimización consta de 4 fases:
 1. Llevar la consulta a una representación interna
 2. Aplicar a la representación interna una serie de transformaciones lógicas
 3. Convertir la consulta transformada en un conjunto de planes de acceso
 4. Elegir el plan de acceso óptimo
- ❑ Las Fases 1 y 2 son independientes de los datos, al contrario que las fases 3 y 4.
- ❑ Es necesario mantener información de la estructura de la base de datos e información estadística sobre su contenido.

Fase 1. Representación interna

- ❑ Elimina aspectos específicos del lenguaje que se está empleando y prepara la consulta para las siguientes etapas del proceso.
- ❑ Suele representarse la consulta a través de un árbol sintáctico.



Fase 2. Transformaciones Lógicas

- ❑ **Objetivo:** Construir expresiones semánticamente equivalentes a la expresión de consulta, con el fin de:
 - eliminar redundancias, evitando esfuerzos duplicados.
 - facilitar los siguientes pasos en el proceso de normalización.
- ❑ En concreto, los objetivos perseguidos por la transformación lógica son:
 - *Estandarización.* Definir un punto de partida estándar para que el optimizador sea portable a distintos sistemas.
 - *Simplificación.* Encontrar al menos una expresión no redundante, semánticamente equivalente a la de entrada.
 - *Preparación.* Seleccionar la expresión simplificada con mejor rendimiento.

Fase 3. Evaluación de los componentes de una consulta

- Hace referencia al proceso de extracción de las tuplas de acuerdo a los criterios de selección.
- Existen distintas estrategias dirigidas a reducir el número de evaluaciones. Por ejemplo, en la expresión:
Seleccionar Título de LIBRO donde Num_copias > 5 y Idioma='Inglés'
Si para una tupla Num_copias > 5 es falso, no tiene sentido verificar la restricción de idioma. Para decidir qué condición verificar primero, puede recurrirse a las estadísticas de la BD.
- Si la operación implica a dos relaciones pueden tenerse en cuenta otros heurísticos. Ejemplo: si se recorren ambas relaciones en un bucle, el bucle externo debe recorrer la relación con menor número de tuplas.

Fase 4. Plan de acceso

- El procedimiento para obtener el plan de acceso será el siguiente:
 - Generar todos los planes de acceso razonables.
 - Incluir en cada plan detalles de la representación física de los datos, ordenaciones, información estadística, etc.
 - Escoger el plan más eficiente. El problema radica en cómo estimar el coste de cierto plan.
- TODO el proceso de optimización (las 4 fases) debe realizarse en un tiempo rentable respecto a la mejora de rendimiento a obtener.