



Diseño de Bases de Datos

(TEMAS 1 Y 2)

- Se estudiará el proceso de diseño de una base de datos dentro del marco más amplio que es la **creación y puesta en marcha de una BD.**
- Se examinarán las **entradas y salidas del proceso de diseño.**
- Veremos el Diseño de Bases de Datos (DBD) en el marco de Diseño del Sistema de Información (SI) :
⇒ ***Integración DBD/SI***
- Presentaremos una metodología para el diseño de BD
⇒ conjunto de modelos y herramientas que nos permiten pasar de una etapa a la siguiente en el proceso de diseño.
- Las **fases** en que se puede dividir el proceso de diseño de una base de datos según la metodología propuesta.

Proceso de creación de una BD.

- ESTUDIO PREVIO Y PLAN DE TRABAJO
 - ⇒ Concretar la voluntad de los directivos de abordar el proyecto, definiendo los objetivos generales de dicho proyecto.
 - ⇒ Evaluación de los medios y costes de la puesta en marcha.
 - ⇒ Definición de la unidad administrativa que tendrá la responsabilidad de la gestión y control de la BD.
 - ⇒ Plan de trabajo detallado con los plazos y medios requeridos.

- CONCEPCIÓN Y SELECCIÓN DEL EQUIPO
 - ⇒ Análisis del sistema existente y necesidades de los usuarios.
 - ⇒ Evaluar las exigencias en cuanto a equipo, especialmente respecto al SGBD y al dimensionamiento del ordenador.

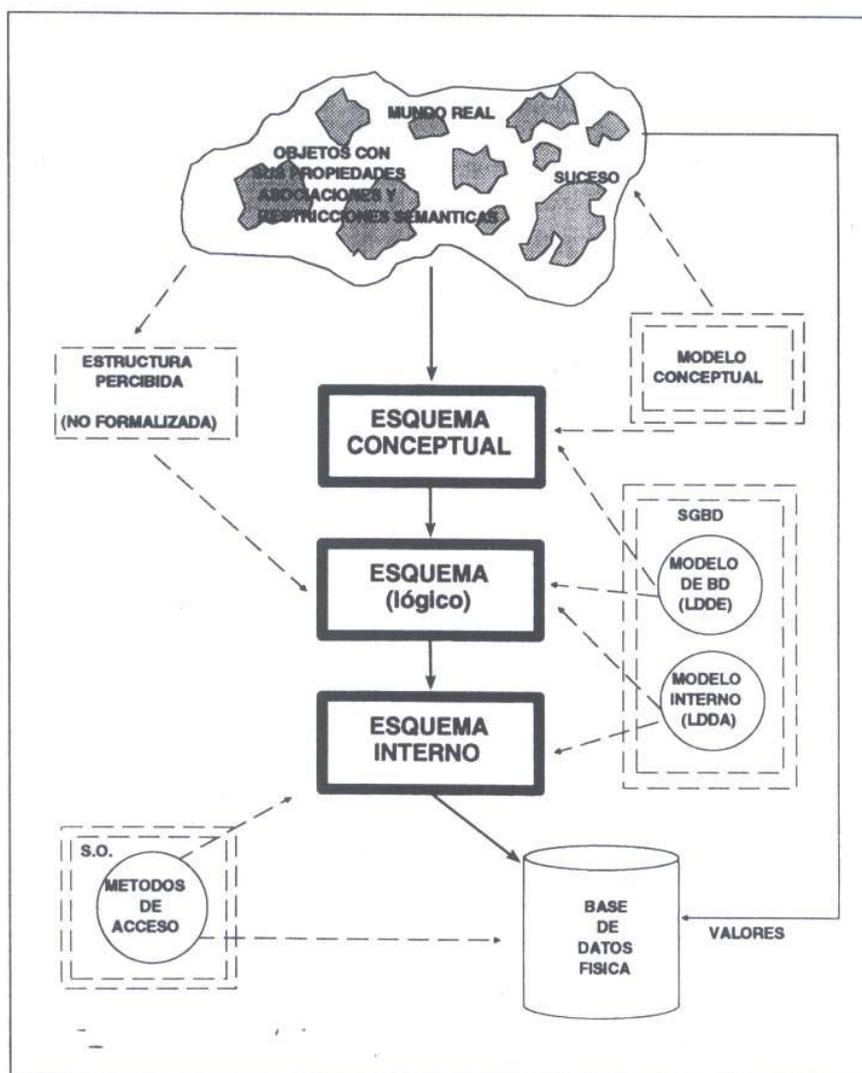
- DISEÑO Y CARGA
 - ⇒ Definición de la estructura teniendo en cuenta el SGBD y el modelo de datos

- PRODUCCION

El proceso de diseño de la BD.

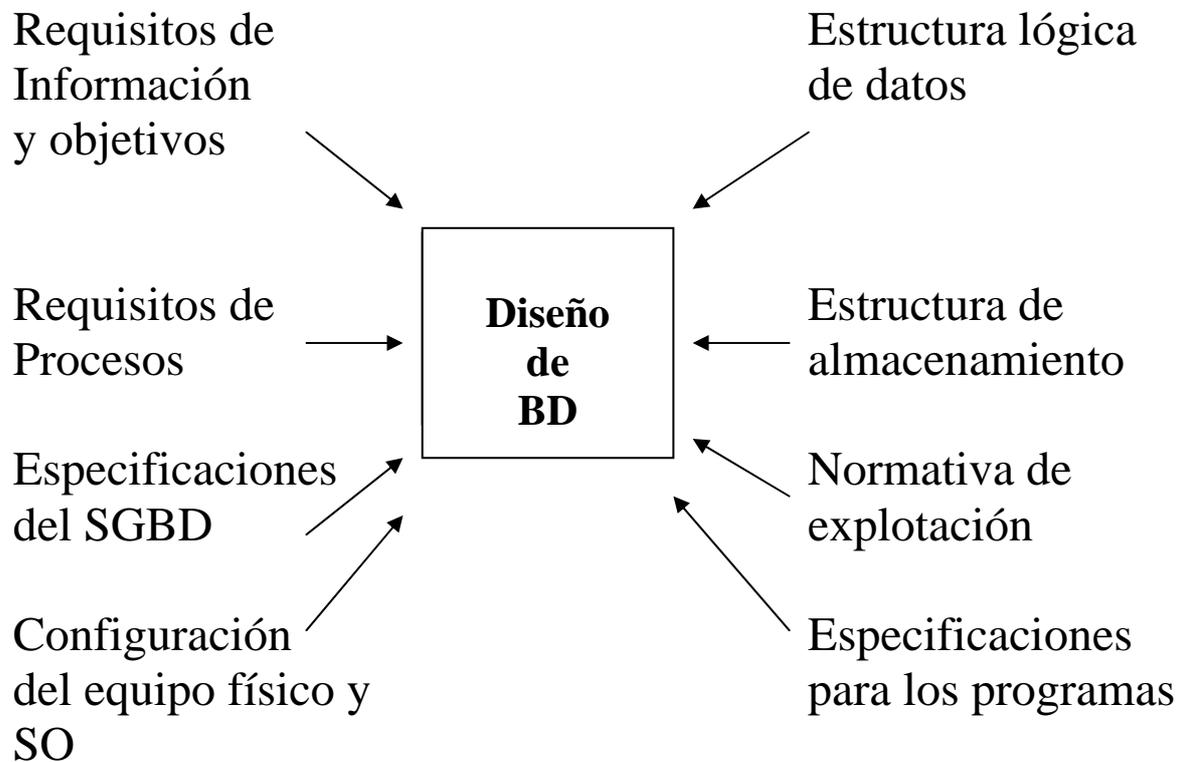
Está muy influenciado por la evolución de:

- La Arquitectura de BD.
 - ⇒ Marca las fases de la metodología.
- Los Modelos de Datos
 - ⇒ Marca los instrumentos para llevar a cabo el diseño.



Detalle de la transformación del mundo real a la BD física

Entradas y Salidas del proceso de diseño.



Entradas y Salidas del proceso de diseño.

Entradas.

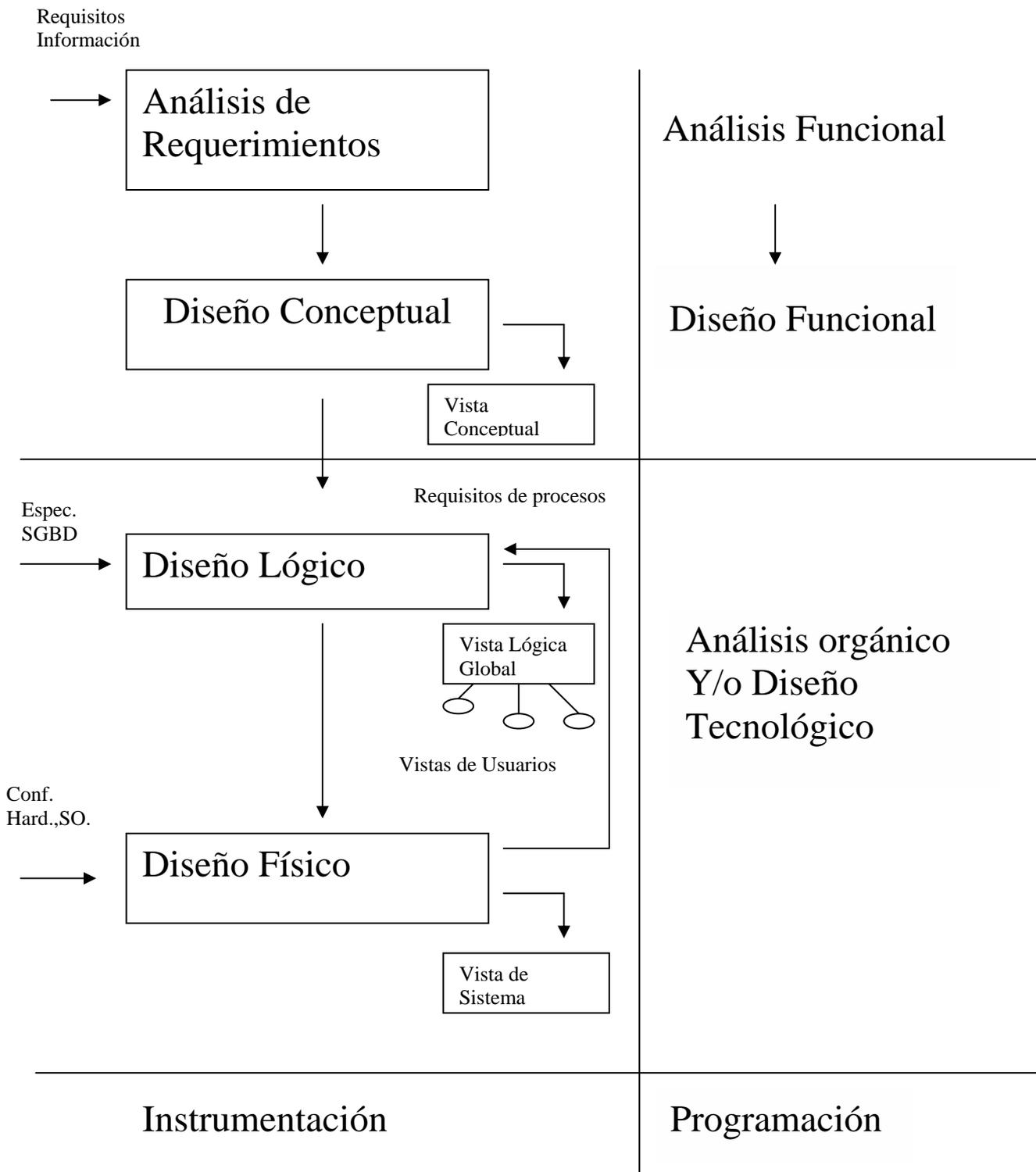
- Requisitos de información y objetivos.
Entrevistas con los usuarios
Análisis de los documentos a generar
Objetivos de la organización.
- Requisitos de procesos.
Características que deben cumplir las aplicaciones,
por ej. Tiempo de respuesta.
- Especificaciones del SGBD
Modelo de datos soportado
Características de rendimiento, seguridad, lenguajes...
- Configuración del equipo físico y del SO
Influirán en el diseño físico y ajuste de la BD.

Salidas.

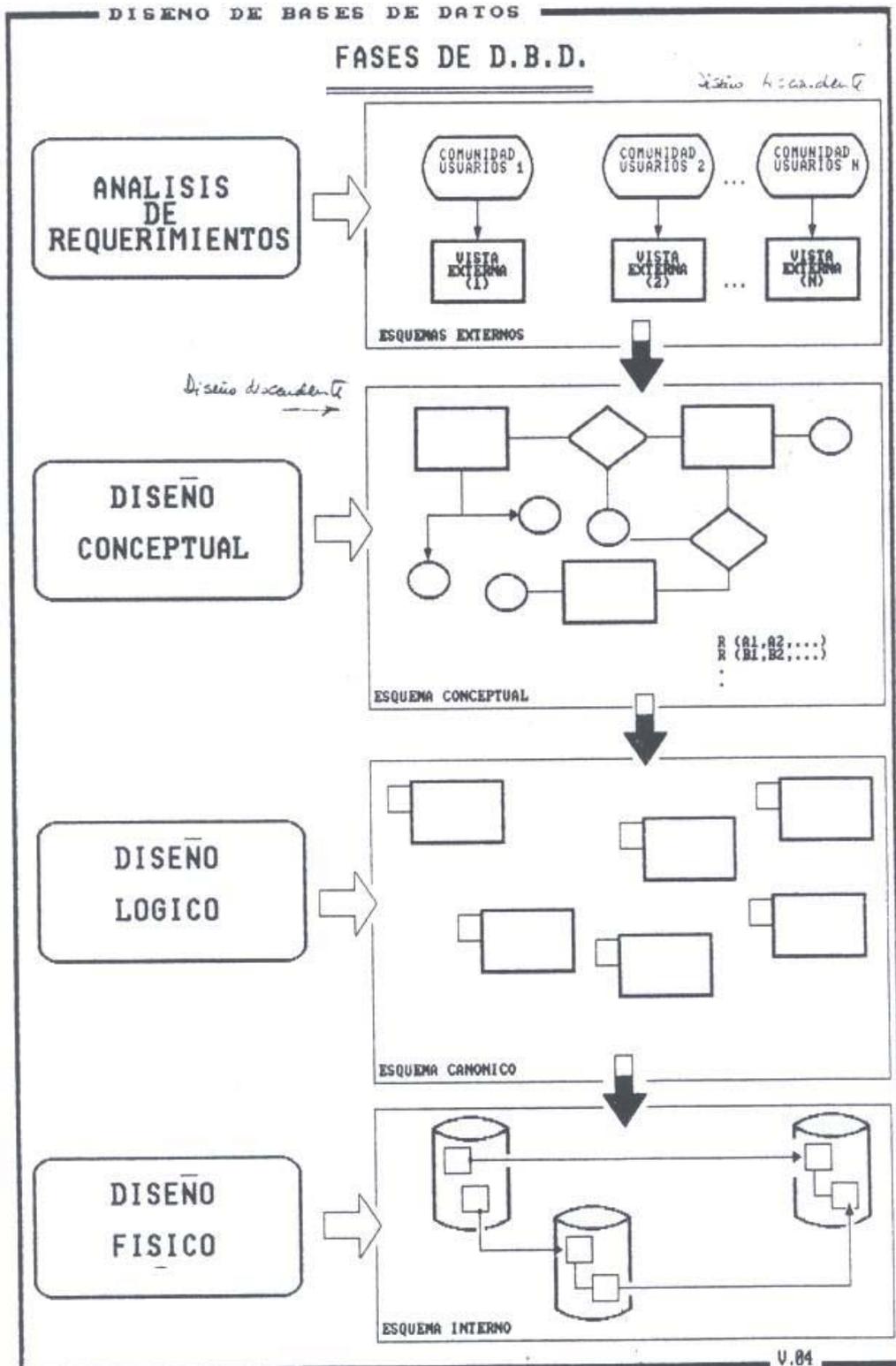
- Estructura lógica de datos.
Esquema conceptual
Esquema lógico en el modelo soportado por el SGBD
Vistas de usuario.
- Estructura de almacenamiento.
Esquema interno con especificaciones físicas como
particiones, definiciones de espacio, índices, cluster..
- Normativa de explotación.
Confidencialidad, seguridad e integridad para la BD.
- Especificaciones para los programas.
Características que no pueden ser recogidas en el
esquema.

Integración DBD/SI

- Realizado conjuntamente.
- Ofreciendo una visión única del proceso
- Cada vez más \Rightarrow Enfoque hacia los datos



Fases de diseño de la BD.



Fases de diseño de la BD.

1. ANÁLISIS DE REQUERIMIENTOS:

- ◆ Captar los requisitos de información de los distintos grupos de usuarios.
- ◆ Información sobre el uso que se piensa dar a la BD.
- ◆ Captar requerimientos operativos
 - Transacciones (críticas y no críticas)

2. DISEÑO CONCEPTUAL:

- ◆ Obtener una buena representación de los recursos de información de la empresa, con independencia de usuarios o aplicaciones en particular, y fuera de consideraciones sobre eficiencia del ordenador.

3. DISEÑO LÓGICO:

- ◆ Transformar el esquema conceptual obtenido en la etapa anterior, adaptándolo al modelo de datos en el que se apoya el SGBD que se va a utilizar .

4. DISEÑO FÍSICO:

- ◆ Trata de conseguir una instrumentación lo más eficiente posible, del esquema lógico.

Análisis de requerimientos

- Es la etapa de percepción, identificación y descripción del mundo real a analizar.
- Se responde a la pregunta ¿ qué representar?.
- Es necesario identificar los usuarios y aplicaciones que van a interactuar con el sistema.

⇒ Identificación de usuarios responsables:

- ◆ Alta dirección.
 - Objetivos y metas corporativas.
 - Visión de funciones importantes
 - Evolución futura.
 - Establecer prioridades.
- ◆ Mandos intermedios.
 - Objetivos detallados.
 - Identificar usuarios intermedios
- ◆ Usuarios operacionales.
 - Requerimientos detallados.
 - Procedimientos.
 - Informes.
 - Formularios..

Análisis de requerimientos

- Método de captura de datos
 - *Entrevistas con los usuarios* de distintos niveles de la organización.
 - Análisis de la documentación existente.
 - Estudio de las reglas de la empresa.
 - Análisis de las transacciones y su frecuencia.
- ⇒ Se obtienen especificaciones de requerimientos mal estructuradas e informales que posteriormente se formalizarán mediante técnicas de especificación de requerimientos (ERD no refinados, DFD..).

Diseño Conceptual

- Se transforma el esquema descriptivo refinándose y estructurándose adecuadamente.
- Se propone el modelo entidad relación para representar el esquema conceptual (EC).
- Diseñar las características de las transacciones con independencia del SGBD.

⇒ ENFOQUES PARA EL DISEÑO CONCEPTUAL

1. Enfoque centralizado (Top-Down).

- ◆ Combinar los requerimientos de los distintos grupos (DBA) antes de iniciar el diseño del EC.
 - ⇒ Sobre dicho esquema se especificarán los esquemas externos (EE)

2. Enfoque de integración de vistas (Bottom-Up)

- ◆ Se crea un esquema para cada grupo de usuarios
 - ⇒ *Integración de vistas* (DBA)
 - ⇒ Las vistas individuales serán los esquemas externos

Diseño Conceptual

⇒ INTEGRACIÓN DE VISTAS (Esquemas)

- ◆ Resolución de conflictos.
- ◆ Análisis de redundancias.

1. Conflictos de nombres

- Homónimos
- Sinónimos
 - Diccionario de datos.
 - Cambiar el nombre a la entidad ó interrelación.

2. Conflictos entre tipos de objetos

- Se transformará el que más convenga.

3. Conflictos de dominios

4. Conflicto entre entidades.

- Una entidad es un subconjunto de otra.
 - La solución es introducir un subtipo.
- Una entidad disjunta con otra y atributos comunes.
 - La solución es crear un supertipo.

5. Conflicto de cardinalidades en interrelaciones.

- ¿Las dos interrelaciones son la misma?

Diseño Lógico

◆ **Diseño lógico estándar.**

⇒ Será necesario haber elegido el modelo de datos, pero sin restricciones ligadas a ningún producto comercial.

◆ **Diseño lógico específico.**

⇒ Se elabora el esquema lógico específico (ELE) que será descrito en el lenguaje de definición de datos del producto comercial elegido.

⇒ ¿En que grado soporta el SGBD los conceptos del modelo lógico estándar?

- Dominios
- Claves primarias
- Claves ajenas
- Otros conceptos

Diseño Físico

- ◆ Objetivos
 - Disminuir los tiempos de respuesta
 - Minimizar el espacio de almacenamiento
 - Optimizar el consumo de recursos

- ◆ Entre las opciones que ofrece el SGBD, se elegirán las estructuras de almacenamiento y caminos de acceso más adecuadas para conseguir un buen rendimiento

- ◆ Después de la implementación y observando el rendimiento del sistema, puede ser necesario afinar el diseño.
 - ⇒ *Indices secundarios.*

- ◆ Las transacciones con restricciones de tiempo importantes se tendrán en cuenta para asignar prioridades.

- ◆ El rendimiento dependerá del tamaño y número de registros de los archivos, será necesario estimar estos parámetros.

- ◆ Tener en cuenta las estimaciones de crecimiento de los archivos.

Diseño Físico

CONSIDERACIONES PARA SGBD RELACIONALES.

1. Para mejorar el JOIN

- ❖ Cuando haya dos relaciones con interrelación 1:N, se puede almacenar como un solo archivo jerárquico, donde se almacena cada registro del lado 1 seguido de los registros del lado N.

- ❖ Desnormalizar las relaciones
 - Repitiendo el atributo en el archivo que se necesite.
 - Almacenar físicamente un archivo que es join de dos archivos.

⇒ *Anomalías de actualización*

Diseño Físico

2. Organización de archivos

- ❖ Se pueden mantener los archivos sin ordenar y crear solamente índices secundarios.
- ❖ Especificar un atributo de ordenación especificando un índice primario (unique) o de agrupamiento (cluster).

3. Definición de índices

- ❖ Para cada atributo(s) que se use con frecuencia en criterios de selección u operaciones de join.
- ❖ Para los archivos que se actualicen con frecuencia, reducir al mínimo la definición de índices.

Modelo de evaluación de un esquema

Sea una relación **R** del esquema $\rho \{R_1, R_2.. R_n\}$

Tuplas(R)	Número de tuplas de R
Valor (R, A_i)	Número de entradas de índices (valores distintos) para el atributo A _i de la relación R
I(R)	Número de índices.
F_I (R)	Frecuencia inserciones de tuplas en R por unidad tiempo
F_D(R)	Frecuencia borrado de tuplas en R.
F_U(R)	Frecuencia modific. de tuplas en R
F_S(R,A_i)	Frecuencia de consultas sobre el atributo A _i
F_J (R,R')	Frecuencia de JOIN entre R y R'

❖ *Suponiendo que el coste de las operaciones es lineal y proporcional al número de índices.*

1. Actualizaciones

(coste por unidad de tiempo):

$$\text{Inserciones} \quad \text{Coste}_I(R) = \lambda_I * I(R) + \mu_I$$

$$\text{Modificaciones} \quad \text{Coste}_D(R) = \lambda_D * I(R) + \mu_D$$

$$\text{Eliminaciones} \quad \text{Coste}_U(R) = \lambda_U * I(R) + \mu_U$$

Modelo de evaluación de un esquema

2. Selecciones:

$$\text{Coste}_S(R, A_i) = \text{IND}(R, A_i) \cdot \lambda_{SS} \cdot \text{Tuplas}(R) + (1 - \text{IND}(R, A_i)) \cdot \lambda_{SI} \cdot (\text{Tuplas}(R) / \text{Valor}(R, A_i))$$

- Coste exploración secuencial
 $\text{IND}(R, A_i) \cdot \lambda_{SS} \cdot \text{Tuplas}(R)$
- Coste exploración indexada
 $(1 - \text{IND}(R, A_i)) \cdot \lambda_{SI} \cdot (\text{Tuplas}(R) / \text{Valor}(R, A_i))$

$\text{IND}(R, A_i) = 0$	Si A_i está indexado
$\text{IND}(R, A_i) = 1$	Caso contrario
λ_{SS}	Índice del coste búsqueda no indexada.
λ_{SI}	Índice del coste de búsqueda indexada.

3. Combinaciones (Joins):

$$\text{Coste}_J(R, R') = \mu_J \cdot \text{Tuplas}(R) \cdot \text{Tuplas}(R')$$

$$\begin{aligned} \text{COSTE GLOBAL} = & \sum_R \{ F_I(R) * \text{Coste}_I(R) + \\ & F_D(R) * \text{Coste}_D(R) + \\ & F_U(R) * \text{Coste}_U(R) + \\ & F_J(R, R') * \text{Coste}_J(R, R') \} + \\ & \sum_A \{ F_S(R, A) * \text{Coste}_S(R, A_i) \} \end{aligned}$$

Modelo de evaluación de un esquema

SOLUCIÓN:

Simulación sobre la formula del coste global o sobre transacciones críticas.

El problema **Mín(COSTE GLOBAL)** se resuelve:

- Diseño lógico \Rightarrow Elección de relaciones

*SOBRENORMALIZACIÓN
DENORMALIZACIÓN*

Contribución de Coste_j

- Diseño físico \Rightarrow Elección índices
 \Rightarrow Criterios acceso

Contribución de Coste_s