

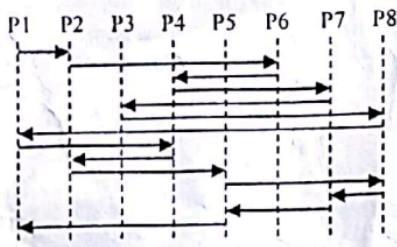
TAA ORDENADO

UNIDAD 11 ALMACENAMIENTO Y ESTRUCTURA DE ARCHIVOS

1. En qué consiste la estrategia RAID y detalle los Niveles 0, 1, 5, 1+0, 0+1
2. Los enfoques o estrategias de organización de registros en archivos de datos en un SGBD..
3. Las medidas de rendimiento de los discos magnéticos. (6p)
4. Los enfoques de organización de archivos un SGBD. (6p)
5. FASD

Tema 8 - 10 puntos La siguiente figura que muestra la planificación de lectura del disco en un SGBD que no utiliza el algoritmo del ascensor. Se pide:

- a- Explique el algoritmo del ascensor y el porqué de su utilidad.
- b- Muestre una planificación basada en el uso del algoritmo del ascensor suponiendo que:
 1. El recorrido del peine empieza en la pista 1,
 2. Se pueden atender varias solicitudes por parada.
 3. Las solicitudes llegan cada 1 ms.
 4. Las lecturas demoran 0,5 ms por solicitud.
 5. El peine se mueve a una razón de 2 pistas/ms
- c- Hallar la tasa de atención de solicitudes por unidad de tiempo sin y con la aplicación del algoritmo del ascensor.



6. Explique detalladamente las medidas de rendimiento que deben ser consideradas respecto a los discos magnéticos.
7. Explique detalladamente las ventajas de configurar un esquema de almacenamiento basada en RAID. Ejemplifique además los niveles 1 y 5.
8. Detalle apropiadamente como se implementa la organización física de bloques denominada estructura de páginas por ranuras para el almacenamiento de registros.
9. Explique las formas de organización física de los archivos de datos en un SGBD.

Los archivos se pueden organizar de cuatro maneras

- **Montículo (Heap):**

- Los registros se colocan en cualquier espacio disponible del archivo, sin seguir un orden específico.
- No están ordenados.
- Son útiles para consultas aleatorias.

- **Secuencial:**

- Los registros se almacenan en orden secuencial, según el valor de su clave de búsqueda.
- Son eficientes para consultas de rango.

- **Hash:**

- Se utiliza una función hash para determinar el bloque donde se almacenará cada registro, basándose en un atributo específico.
- Permite ubicar rápidamente los registros al ingresar nuevos datos.
- Es adecuado para consultas unitarias (que buscan una sola fila).

- **Agrupación:**

- Almacena datos de diferentes relaciones en un mismo archivo.
- Los registros relacionados se guardan en el mismo bloque.
- Facilita las consultas de tipo "join".

Nota: La organización secuencial es la única que siempre funcionará, independientemente del tipo de consulta.

10. **Medidas de rendimiento para la elección de unidades de discos magnéticos:**

- **Tiempo de acceso:**

- Tiempo que tarda en leer o escribir datos desde la petición hasta la confirmación.

- **Tiempo de búsqueda:**

- Tiempo que tarda el cabezal en ubicarse en la pista correcta.
- Los discos más grandes pueden tardar más tiempo en encontrar la pista deseada, mientras que los discos más pequeños pueden ser más rápidos.

- **Latencia rotacional:**

- Tiempo que tarda el sector deseado en llegar a la posición de lectura/escritura una vez que el cabezal está en la pista correcta.

- **Tiempo de transferencia:**

- Volumen de datos transferidos por unidad de tiempo, es decir, la velocidad de recuperación o almacenamiento de la información.

- **Tiempo medio entre fallos (MTBF):**

- Tiempo promedio de vida útil del disco, indicando cuándo es probable que falle debido a problemas acumulados.

Criterio sobre la medida más determinante:

- "A mi criterio sería la latencia rotacional, que dependiendo de donde está el dato en el disco este puede llegar a tardar más o menos. Por ej puede llegar a completar una vuelta para encontrar la información o como puede encontrarla al instante."

12. Explique resumidamente y en sus términos, la organización de un bloque de datos para registros de longitud variable.

Son utilizados en BD para cuando se requiere guardar múltiples tipos de registros en una misma tabla, estos registros pueden tener una longitud variable o campos repetidos. Por lo que se puede decir que se adaptan al tamaño de la información incluida en ellos, en estos se utilizan delimitadores tanto para el fin campo como para el fin de registro.

Estos pueden ser implementados como archivos de registros de longitud fija mediante los siguientes esquemas, Espacio reservado, Representación con listas, Estructura de bloques anclas y de desbordamiento.

13. Explique resumidamente y en sus propios términos las posibles formas organización de los registros en archivos para la implementación de archivos de tablas/datos.

- Registros de longitud fija: Cada registro posee una longitud igual a los demás y se insertan al final.

- Lista libre: Cada registro tiene una cabecera sin datos pero que apunta a la siguiente cabecera vacía.
- Registros de longitud variable: Cada registro posee longitud variable y en general un espacio libre disponible.
- Archivos secuenciales: Registros se introducen uno detrás del otro.
- Organización en agrupación: Se juntan varias columnas que se suelen solicitar en conjunto para recuperarlas más rápido.

14. Explique el algoritmo del Ascensor implementado en para el acceso a la información en Unidades de Discos Magnéticos.

Planificación de Brazos del Disco.

■ Algoritmos que se encargan de planificar el movimiento del peine para maximizar la cantidad de solicitudes atendidas por unidad de tiempo.

■ Basados en el algoritmo del ascensor: mover el peine en una dirección hasta atender todas las solicitudes en dicha dirección y luego repetir la operación en dirección inversa.

■ Se implementan en el controlador de disco.

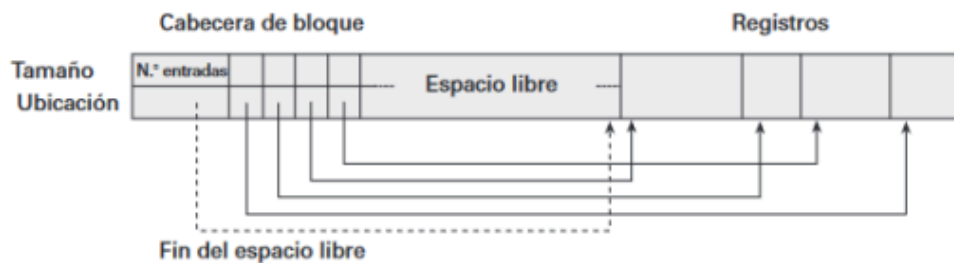
15. Indique además las principales métricas de rendimiento para la selección de unidades de almacenamiento.

- a. Tiempo de Acceso: Tiempo que toma leer o escribir datos desde su petición hasta su confirmación.
- b. Tiempo de búsqueda: Tiempo que se tarda en ubicar el cabezal de lectura en la pista correcta. 4 a 10 ms.
- c. Latencia Rotacional: Tiempo que tarda en aparecer el sector a escribir/leer debajo de los cabezales. 4 a 15 ms.
- d. Tasa de Transferencia: Volumen de datos por unidad de tiempo que se pueden guardar o recuperar del disco. Típicamente de 30 a 100 MB/s para lectura 20 a 85 MB/s para escritura.
- e. Tiempo medio de fallo (MTBF):
El lapso de tiempo que un solo disco puede trabajar continuamente sin fallas. De 3 a 5 años

16. Explique la importancia del Almacenamiento Redundante (RAID)

La principal razón de usar raid es obviamente mantener los datos seguros en caso de fallas de disco, alargando el tiempo de probabilidad de que los discos fallen al mismo tiempo(el que guarda los datos y su respaldo). Teniendo así un respaldo seguro y completo de la información guardada

17. Ilustre y describa la Estructura de Páginas con Ranura para la implementación de Registros de Datos de Longitud Variable.



Hay una cabecera al principio de cada bloque que contiene la información siguiente:

1. El número de elementos del registro de la cabecera
2. El final del espacio vacío del bloque
3. Un array cuyas entradas contienen la ubicación y el tamaño de cada registro

Los registros se ubican de manera contigua en el bloque, empezando desde el final del mismo. El espacio libre dentro del bloque es contiguo, entre la última entrada del array de la cabecera y el primer registro. Si se inserta un registro se le asigna espacio al final del espacio libre y se añade a la cabecera una entrada que contiene su tamaño y su ubicación.

18. Detalle apropiadamente como se implementa la organización física de bloques denominada estructura de páginas por ranuras para el almacenamiento de registros.

Existen diferentes técnicas para implementar los registros de longitud variable.

La estructura de páginas con ranuras se utiliza habitualmente para organizar los registros en bloques, y puede verse en la Figura 11.9. Hay una cabecera al principio de cada bloque, que contiene la información siguiente:

1. El número de elementos del registro de la cabecera.
2. El final del espacio vacío del bloque.
3. Un *array* cuyas entradas contienen la ubicación y el tamaño de cada registro.

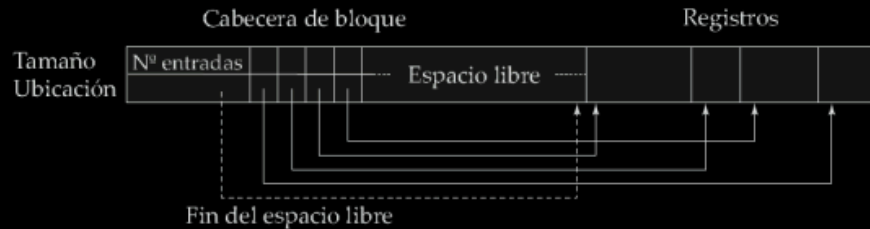


Figura 11.9 Estructura de páginas con ranuras.

El método de la página con ranuras se usa mucho para manejar los registros de longitud variable en los bloques de disco.

19. Explique la estructura de "páginas con ranuras" para la implementación de bloques de registros de longitud variable en un SGBD.

La estructura puede describirse de la siguiente manera

-Una cabecera al principio de cada bloque que contiene la información siguiente

-El número de elementos del registro de la cabecera

-El final del espacio vacío del bloque

-Un array cuyas entradas contienen la ubicación y el tamaño de cada registro

Los registros se ubican de manera contigua en el bloque, empezando desde el final del mismo. El espacio libre dentro del bloque es contiguo, entre la última entrada del array de la cabecera y el primer registro. Si se inserta un registro se le asigna espacio al final del espacio libre y se añade a la cabecera una entrada que contiene su tamaño y su ubicación.

20. Explique brevemente la ventaja de los Arreglos Redundantes de Discos (RAID) en cuanto a: 1-) Rendimiento. 2-) Fiabilidad.

La principal razón de usar raid es obviamente mantener los datos seguros en caso de fallas de disco, alargando el tiempo de probabilidad de que los discos fallen al mismo tiempo(el que guarda los datos y su respaldo). Teniendo así un respaldo seguro y completo de la información guardada.

En cuanto al rendimiento se puede decir que dependiendo el tipo de raid que

se usa se puede mejorar el tiempo de lectura de datos o de escritura. Ya que la misma información se divide y se podrá guardar en varios discos acortando estos tiempos.

21. Indique la importancia del Catálogo del Sistema/Diccionario de Datos.

(ii) Detalle mínimamente la información almacenada en el mismo

a. Guarda información acerca del sistema y sus datos (metadatos) y los guarda en la misma

base de datos, se debe mantener en memoria para permitir un acceso rápido al mismo.

i. Información acerca de las relaciones: Nombre de las relaciones, Nombre y tipo de los atributos de cada relación, Nombre y definiciones de las vistas y las restricciones de integridad.

ii. Cuentas de usuario incluyendo nombre y contraseña

iii. Información estadística y descriptiva: Número de tuplas de cada relación y Método de almacenamiento de cada relación.

iv. Información de la organización física: Tipo de almacenamiento para cada relación y Localización física de cada relación

v. Información de los índices: Nombre del índice y de su relación, Tipo de índice y atributos sobre los cuales se crea el índice

22. **Suponga que se tiene un SGBD con soporte de medio de almacenamiento basado en un disco magnético de 200 pistas en el que: (1) actualmente el cabezal de lectura/escritura se encuentra en la pista 50 con dirección al extremo externo del disco y que, (2) se debe atender la siguiente serie de solicitudes en las siguientes pistas: 82, 170, 43, 140, 24, 16, 190, 75, 88, 3, 200, 55, 130, 60. Conforme a lo anterior estímesese:**

a. El número de pistas procesadas asumiendo que no se aplica el algoritmo del ascensor.

b. El número de pistas procesadas asumiendo la ampliación del algoritmo del ascensor.

UNIDAD 12 INDICES

1. Los tipos de índices estudiados conforme su organización física y cuándo conviene aplicarlos.

- **Índice Ordenado:** Los registros índices están ordenados de acuerdo al valor de la clave de búsqueda. Pueden ser primarios o secundarios y conviene utilizarlos cuando se tiene búsquedas en rangos .

Primario: Son aquellos cuya clave especifica el orden secuencial del archivo indexado, ordena físicamente la tabla

Secundario: Son aquellos índices cuya clave de búsqueda no determinan el orden secuencial del archivo ordenado, no ordena físicamente la tabla

- **Índice asociativos o hash:** Los registros índices están almacenados en bloques de acuerdo a una función hash, útiles en condiciones de igualdad.
- **Índice Bitmap:** Los índices bitmaps son un tipo especial de índices diseñados para consultar eficientemente en múltiples claves y son útiles tanto para consultas de sobre un atributo o sobre varios atributos, siempre y cuando exista un índice bitmap en cada atributo.

2. Qué características de los Árboles B y B+ los hace apropiados como estructura para implementar índices ordenados.

La estructura de los árboles B y B+ y su funcionamiento, es decir, la forma en que se añaden y eliminan los datos es conveniente para los índices ordenados, pues al hacer estas operaciones (insert y delete) el árbol se queda ordenado, esto ayuda a determinar si el dato también existe o no y retornarlo en caso de éxito en un tiempo casi lineal. Además esta estructura optimiza la búsqueda, porque los datos siempre están en la rama en donde se empieza a buscar y no en las otras

3. La siguiente tabla corresponde al estado actual del archivo de la relación cliente, en el que cada bloque del archivo se corresponde con una fila, se pide:

- a. **Construir un índice en forma de árbol B+ con nodos de 4 punteros para la clave primaria id, suponiendo que los registros/filas fueron**

insertados según el orden alfabético de la columna nombre.

- b. Construir un índice hash estático cerrado con cajones de 3 elementos cuya función de asociación es " $x \bmod 6$ " sobre la columna saldo, siendo x el valor de cada fila en dicha columna.
- c. Explique en cada caso anterior, si el índice es primario o secundario, denso o disperso y el porqué en cada caso.

id	nombre	saldo
1	Preston Schwartz	282
2	Cathleen Steele	159
3	Tatyana Russo	367
4	Libby Madden	431
5	Orla Reid	317
6	Vivian Cherry	361
7	Kirk Jensen	317
8	Amanda Macias	190
9	Barry Morris	338
10	Lee López	437
11	Elliott Fowler	367
12	Paula Johns	190

4. Los tres tipos de índices estudiados y cuando conviene aplicarlos. (6p)
5. La siguiente tabla corresponde al archivo de la relación cliente, en el que cada bloque del archivo corresponde a 1 fila, se pide:
 - a. Construir un índice en forma de árbol B+ con nodos de 4 punteros para la clave primaria id, suponiendo que los registros fueron insertados según el orden de los valores de la columna nombre.
 - b. Construir un índice hash estático cerrado con cajones de 3 elementos cuya función de asociación es " $x \bmod 5$ " sobre la columna saldo, siendo x el valor de cada fila en dicha columna.
 - c. Explique a donde apuntan los registros índices para caso anterior.

id	nombre	saldo
1	Preston Schwartz	282
2	Cathleen Steele	159
3	Tatyana Russo	367
4	Libby Madden	431
5	Orla Reid	317
6	Vivian Cherry	361
7	Kirk Jensen	317
8	Amanda Macias	190
9	Barry Morris	338
10	Lee Lopez	437
11	Elliott Fowler	367
12	Paula Johns	190

6. Ilustre la construcción de un índice Hash basado en la técnica de Asociación Extensible para la columna nombre-sucursal considerando para ello la tabla y los valores de hash indicados a continuación.

C-217	Barcelona	750
C-101	Daimiel	500
C-110	Daimiel	600
C-215	Madrid	700
C-102	Pamplona	800
C-201	Pamplona	900
C-218	Pamplona	700
C-227	Reus	700
C-305	Ronda	300

nombre-sucursal	nombre-sucursal
Barcelona	0010 1101 1111 1011 0010 1100 0011 0000
Daimiel	1010 0011 1010 0000 1100 0110 1001 1111
Madrid	1100 0111 1110 1101 1011 1111 0011 1010
Pamplona	1111 0001 0010 0100 1001 0011 0110 1101
Reus	0011 0101 1010 0110 1100 1001 1110 1011
Ronda	1101 1000 0011 1111 1001 1100 0000 0001

7. La siguiente tabla corresponde al estado actual del archivo de la relación Cliente, en el que cada bloque del archivo corresponde a 1 fila, se pide:
- Construir un índice en forma de árbol B+ con nodos de 4 punteros para la clave primaria id, suponiendo que los registros/filas fueron insertados según el orden alfabético de la columna nombre.
 - Construir un índice hash estático cerrado con cajones de 4 elementos cuya función de asociación es " $x \bmod 4$ " sobre la columna saldo, siendo x el valor de cada fila en dicha columna.

- c. Explique detalladamente en cada caso si el índice es primario o secundario.

Id	nombre	saldo
1	Preston Schwartz	282
2	Cathleen Steele	159
3	Tatyana Russo	367
4	Libby Madden	431
5	Orla Reid	317
6	Vivian Cherry	361
7	Kirk Jensen	317
8	Amanda Macias	190
9	Barry Morris	338
10	Lee Lopez	437
11	Elliott Fowler	367
12	Paula Johns	190

8. FAS

Conforme la siguiente tabla que corresponde al archivo de la relación EVALUACIÓN, en el que cada bloque del archivo corresponde a 1 fila, se pide:

- Construir un índice en forma de árbol B+ con nodos de 4 punteros para la clave primaria ID, suponiendo que los registros fueron ingresados (insertados) según el orden de la columna NOMBRE.
- Construir un índice hash estático cerrado con cajones de 3 elementos cuya función de asociación es " $x \bmod 5$ " sobre la columna PUNTAJE, siendo x el valor de cada fila en dicha columna.
- Explique para cada caso (a y b) anterior, si los registros índices apuntan o no a la tabla.

ID	NOMBRE	PUNTAJE
1	Preston Schwartz	282
2	Cathleen Steele	159
3	Tatyana Russo	367
4	Libby Madden	431
5	Orla Reid	317
6	Vivian Cherry	361
7	Kirk Jensen	317
8	Amanda Macias	190
9	Barry Morris	338
10	Lee Lopez	437
11	Elliott Fowler	367
12	Paula Johns	190

9. Explique en sus términos a que corresponde un índice y a que un archivo índice.

Asímismo enumere al menos tres tipos de índices estudiados.

a. Índices

-Los índices son aquellos que te ahorran tiempo en hacer una búsqueda en una colección basta de registros, por ej un índice de un libro nos puede guiar más rápidamente al cap que estamos buscando.

-Los índices son estructuras de una base de datos utilizados para acelerar el acceso a los datos deseados.

b. Índices ordenados: Son aquellos en donde los registros índice son almacenados según el orden definido por la clave de búsqueda, por lo tanto son archivos ordenados siempre.

c. Índices asociativos: Son los que los registros índice son almacenados en un cajón partir del uso de una función de hash o de asociación.

d. Índices bitmap: Estos índices están diseñados para consultar eficientemente con múltiples claves, siempre y cuando haya un índice bitmap en cada atributo. Este está compuesto por un array de bits para cada valor de atributo del índice.

10. **Explique resumidamente y en sus términos, cuando debería optarse por el uso de índices bitmaps.**

Los índices bitmaps son un tipo de índices apropiados en escenarios en donde la carga de trabajo esté basada consultas sobre atributos de valores discretos o discretizables y asimismo, en escenarios en donde se espera que la misma carga de trabajo requiera de consultar eficientemente en múltiples claves basados en esos atributos

11. **Indique para qué tipo de consultas resulta más apropiado:**

(a) un índice ordenado.

(b) un índice hash.

(c) un índice bitmap.

Una consulta del tipo (múltiples claves) es apropiado para un índice bitmap

Una consulta del tipo (igualdad e igualdad múltiple) es apropiado para un índice hash

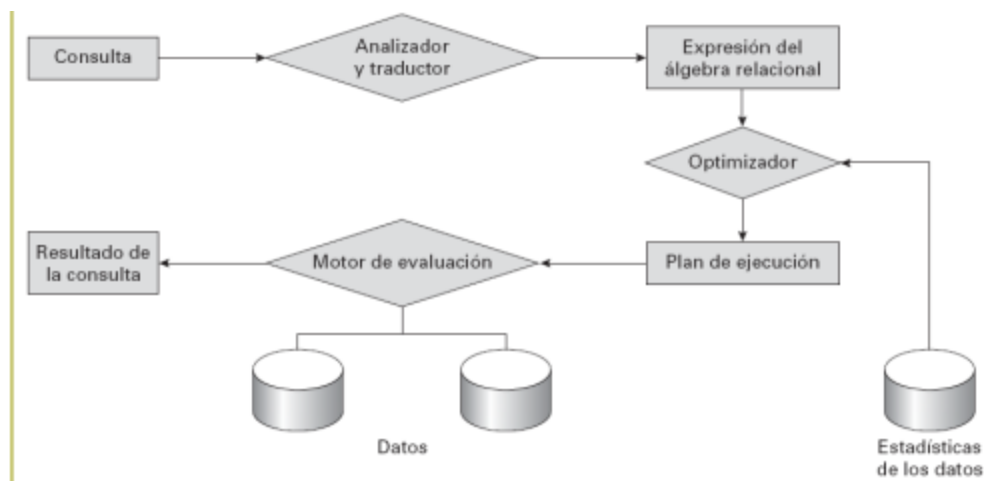
Una consulta del tipo (desigualdad) es apropiado para un índice ordenado

UNIDAD PROCESAMIENTO DE CONSULTAS

1. Los pasos para el Procesamiento de Consultas. Haga el diagrama y explique cada paso.

Los pasos involucrados en el Procesamiento de las consultas son:

- Análisis y Traducción
- Optimización.
- Evaluación.



2. Dada una consulta de selección cualquiera basada en una condición de igualdad simple, indique que condiciones físicas de la BD obligarían a un SGBD utilizar cada uno de los siguientes algoritmos, indique además el costo asintótico correspondiente a cada caso:

a. Selección basada en Búsqueda Lineal.

Se utiliza cuando:

- No se tiene un índice de clave de búsqueda.

- La tabla no está ordenada.

En este caso, es necesario revisar todos los bloques de datos para verificar si cumplen con la condición de selección. Como la búsqueda es secuencial, su costo es:

$$\text{costo} = Br.$$

b. **Búsqueda Binaria.**

Se utiliza cuando:

- No se tiene un índice de clave.
- La tabla está ordenada según la clave de búsqueda.

Este algoritmo es similar a la búsqueda binaria en un array ordenado. Se divide repetidamente el conjunto de bloques hasta encontrar el registro buscado. En el peor de los casos, el costo de búsqueda es:

$$= \log_2(Br) + (CS(A, r) / fr) - 1.$$

c. **Búsqueda en índice primario para un atributo clave.**

Se utiliza cuando:

- Existe un índice primario construido sobre el atributo clave de la relación.

Dado que el índice es único para cada registro, se puede utilizar el índice para acceder directamente al bloque donde se encuentra el registro.

$$\text{El coste} = AA_i + 1.$$

d. **Búsqueda en índice secundario para atributo no clave.**

Se utiliza cuando:

- Se tiene un índice secundario sobre un atributo que no es clave.

Si el atributo indexado es una clave candidata, el índice apunta directamente al registro deseado. Sin embargo, si no es clave, el índice

apunta a múltiples registros y puede ser necesario acceder a varios bloques diferentes.

En el peor caso, cada registro se encuentra en un bloque distinto, lo que implica el siguiente costo:

$$AA_i + 1 + CS(A, r).$$

3. Dada la consulta de abajo, proporcione una traducción inicial de la misma en algebra relacional y luego proceda a ilustrar con la misma al menos dos casos de optimización conforme las reglas de equivalencias estudiadas.

```
select e.LNAME from EMPLEADO e
join TRABAJA_EN te on (te.EMPLEADO = e.ID)
join PROYECTO p on (p.ID = te.PROYECTO)
where p.PNOMBRE = 'AQUARTUS' and e.FECHA_NAC >= '2000-01-01'
```

4. Desarrolle el planteamiento hipotético de cuatro casos basados en una consulta de selección en la forma: `select * from <tabla> where <tabla>.<columna> = <valor>` y que deriven cada uno en una operación selección con coste o estimación de coste diferente.

Dada la consulta `select * from A join B on A.a = B.b`, teniendo en cuenta que:

- La tabla A se encuentra almacenada en 20 bloques,
- La tabla B se encuentra almacenada en 15 bloques,
- La memoria cuenta con actualmente con 10 bloques libres para evaluar la consulta, y
- Que la consulta está planificada para ser evaluada conforme el algoritmo de Bucle Anidado por Bloques,

Calcule cuál será el coste de evaluación de la consulta si:

1. La estrategia de reemplazo de bloques en la memoria es LRU.
2. La estrategia de reemplazo de bloques en la memoria es MRU.

5. Ejemplifique cuatro planteamientos hipotéticos que ilustren cada uno la aplicación de un algoritmo de selección diferente indicando además el coste o estimación de coste del mismo.
6. Dada la consulta `select * from A a join B b on A.a = B.b`, y teniendo en cuenta que:
- La tabla A se encuentra almacenada en 12 bloques de disco.
 - La tabla B se encuentra almacenada en 25 bloques de disco.
 - La memoria cuenta actualmente con 10 bloques libres para evaluar la consulta, y
 - Que la consulta está planificada para ser evaluada conforme el algoritmo de Bucle Anidado por Bloques,

Calcule cuál será el coste de evaluación de la consulta si:

1. La estrategia de reemplazo de bloques en la memoria es LRU.
 2. La estrategia de reemplazo de bloques en la memoria es MRU.
7. **Explique resumidamente, en orden y en sus términos, los pasos que se realizan para el procesamiento de consultas por parte de un SGBD.**
- a. **Análisis y Traducción:**
Verifica la sintaxis. Se revisa en el catálogo del sistema que todos los identificadores sean nombres de objetos de la base de datos. Luego se realiza la traducción a expresiones equivalentes en el álgebra relacional. El resultado es el conjunto o serie de expresiones del álgebra relacional a evaluar para obtener el resultado de la consulta.
 - b. **Optimización:**
Busca la expresión con menor costo, y con eso elabora un plan de ejecución en base a las estadísticas. Devuelve como resultado el plan de ejecución de la consulta.
 - c. **Evaluación:**
Decide si implementar la materialización o Encausamiento.:

d. Materialización:

En la cual se evalúa una operación a la vez. empezando por las expresiones de nivel más bajo del árbol de operadores, el resultado se materializa en un archivo temporal. Se usa el double buffering, para solapar cómputo y escritura.

e. Encausamiento:

Se evalúan varias operaciones simultáneamente. Los resultados de una operación se pasan a la siguiente utilizando una memoria intermedia, sin utilizar archivos temporales. Las entradas y salidas de las operaciones no están disponibles todas a la vez.

8. Dada una consulta de selección basada en una condición de igualdad única y simple (por ejemplo: `select * from persona where numero_ci = 12345`), explique los diferentes algoritmos de selección que podrían ser aplicados por el SGBD y el porqué.

UNIDAD OPTIMIZACION DE CONSULTAS

1. Desarrolle el algoritmo de Programación Dinámica para la optimización del orden de reunión en consultas. Explique detalladamente cuál es el principio detrás del mismo, así como la complejidad computacional espacial y temporal que corresponde al mismo.
2. Explique el costo asintótico espacial y temporal del Algoritmo de Programación Dinámica: (a) sin la optimización de evaluación de árboles en profundidad por la izquierda. (b) con la optimización de evaluación de árboles en profundidad por la izquierda

- La optimización de evaluación de árboles en profundidad por la izquierda reduce significativamente el **costo temporal** del algoritmo de PD, pasando de $O(3^n)$ a $O(2^n)$.
- El **costo espacial** sigue siendo $O(2^n)$, pero la optimización hace que el algoritmo sea más eficiente en términos de tiempo de ejecución, especialmente para consultas con un número moderado de relaciones ($n \leq 10$).

3. Explique la importancia de las reglas de equivalencia del algebra relacional desde el punto de vista de la optimización de consultas. Proporcione y explique al menos dos ejemplos de la aplicación de alguna regla de equivalencia en un caso de optimización.
4. Describa la lógica del Algoritmo de Programación Dinámica para la Optimización de Consultas. Indique los costos asintóticos del Algoritmo tanto en complejidad espacial como temporal

Para buscar el mejor árbol de join de n relaciones n Considerar todos los planes de la forma:

- $S1 (S - S1)$ donde $S1$ es un subconjunto no vacío de S .
- Computar recursivamente el costo de cada subconjunto de reunión de S . Escoger la más barata de las $2^n - 1$ alternativas.
- Caso base de la recursión: acceso a una sola relación. p Utilizar el modo de acceso más apropiado.
- Cuando se computa el plan para cualquier subconjunto, se guarda el resultado para utilizarlo cuando sea requerido nuevamente, en vez de volver a calcularlo.

La complejidad temporal del algoritmo de P.D. es $O(3^n)$
La complejidad espacial es $O(2^n)$

UNIDAD TRANSACCIONES

1. El concepto de transacción y las propiedades de las mismas que debe ser aseguradas por el SGBD (6p.)
2. Explique detalladamente las propiedades de las transacciones que deben ser aseguradas por un SGBD conforme el estándar SQL.
3. SDFA

Tema 6 – 5p – Determine si las siguientes planificaciones son o no secuenciables en cuanto a conflictos e indique cual es el orden de las transacciones en caso afirmativo.

P1	
T1	T2
Leer(A)	Escribir(A)
Escribir(A)	Escribir(A)

Si

P2	
T1	T2
Leer(A)	Escribir(B)
Leer(B)	Escribir(A)

no!

P3		
T1	T2	T3
Escribir(A)	Escribir(B)	Escribir(C)
	Leer(C)	

Si!

T1, T3, T2

4. FASDF

Tema 7 - 5p - Determine si las siguientes planificaciones son o no secuenciables en cuanto a vistas e indique cual es el orden de las transacciones en caso afirmativo.

P1	
T1	T2
Leer(A)	Escribir(A)
Escribir(A)	Escribir(A)

P2	
T1	T2
Leer(A)	Escribir (B) Escribir (A)
Leer(B)	

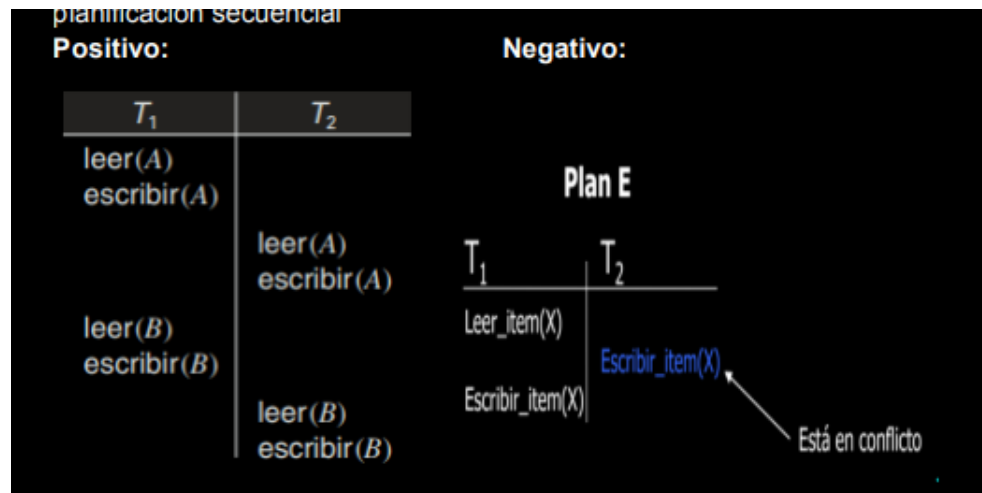
P3		
T1	T2	T3
Leer(A)	Leer(B)	Escribir(B)
Escribir(A)	Escribir(B)	Leer(A) Escribir(B)

τ_1
$\text{free}(A)$
$\text{resizable}(A)$

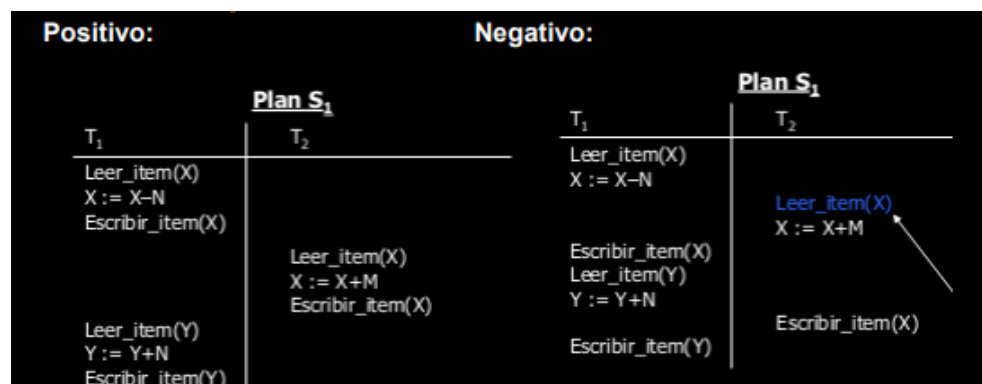
τ_1, τ_2, τ_3

5. Explique en el contexto de los Protocolos del Control de Concurrency que se entiende por: (1) Planificación, (2) Planificación Secuenciable, (3) Secuencialidad en Cuanto a Conflicto y (4) Secuencialidad en Cuanto a Vistas.

- a. Una planificación para un conjunto de transacciones consiste en un ordenamiento de todas las operaciones de dichas transacciones, pero conservando el orden interno en cada transacción individual.
- b. Un plan S de n transacciones es secuenciable si es equivalente a algún plan secuencial de las mismas n transacciones. Un plan secuenciable es siempre correcto. Las planificaciones pueden ser secuenciales en cuanto a vistas o en cuanto a conflictos.
- c. Se dice que existe secuencialidad en cuanto a conflictos cuando dos transacciones realizan operaciones sobre un mismo dato, tal que al menos una de ellas es escribir, entonces, el orden de ambas operaciones es importante. Se dice que una planificación es secuenciable en cuanto a conflictos cuando puede convertirse en una planificación secuencial.



- d. Se dice que hay secuencialidad en cuanto a vistas cuando es posible convertir una transacción T concurrente en una secuencial, cumpliendo los criterios: todas las operaciones "leer" deben leer el mismo dato inicial o producido por una operación escribir previa a la misma, y toda operación "escribir" que haya ocurrido al final de una transacción, debe mantenerse al final de la misma.



6. Defina el concepto de Transacción en SGBDs.

Una transacción es un conjunto de operaciones, que acceden y posiblemente actualizan los datos, y que son tratados como una unidad.

7. Indique y defina las fases del Ciclo de Vida de una Transacción.

a. Activa

Es el estado al que pasa luego de comenzar su ejecución, y no cambia de

estado durante
las operaciones de lectura y escritura de ítems.

b. Parcialmente confirmada

Pasa a este estado cuando acaba (END). En este punto se realizan controles de concurrencia, y en el caso de fallo, los cambios no se graban de forma permanente. Pasa al siguiente estado cuando al hacer commit.

c. Confirmada

Punto de confirmación y conclusión de la ejecución.

d. Fallida

Si uno de los chequeos falla o si es abortada durante su estado activo.

e. Abortada

Se deshacen los efectos de sus operaciones de escritura sobre la BD.

f. Terminada

Pasa a este estado cuando la transacción desaparece del sistema.

8. Indique y defina las propiedades que un SGBD debe asegurar para las Transacciones.

a. Atomicidad

Una transacción es una unidad atómica de procesamiento; es realizada enteramente o no es realizada en nada.

b. Consistencia

La ejecución de la transacción debe pasar la BD desde un estado consistente a otro también consistente.

c. Aislamiento

Una transacción no deberá hacer visible sus modificaciones a otras transacciones hasta que esté confirmada.

d. Persistencia

Cuando una transacción cambia la BD y los cambios son confirmados, estos cambios no deben perderse por fallos posteriores

9. Describa la función del Componente de Gestión de Concurrencia en un SGDB.

Se encarga de asegurar que el efecto de la ejecución concurrente de un conjunto de transacciones que se intercalan en el tiempo sea igual a que las transacciones se hubiesen ejecutado secuencialmente

10. (Describa la estructura de datos utilizada para la gestión y concesión de bloqueos.

Para la gestión de bloqueos se utiliza una tabla de bloqueos que almacena los bloqueos concedidos y las peticiones solicitadas. Esta tabla es una tabla hash en memoria, cuyas entradas corresponden a los elementos de datos bloqueados, y por cada uno elemento de datos hay una lista enlazada con los registros de las solicitudes concedidas pero no liberadas y solicitudes pendientes

11. Describa el protocolo de Bloqueo en 2 Fases. (b) Indique cuales son las variantes del mismo.

a. Describe

Este protocolo exige que cada transacción realice sus peticiones de bloqueo y desbloqueo en dos fases:

■ Fase de crecimiento: La transacción puede obtener bloqueos pero no liberarlos.

■ Fase de decrecimiento: La transacción puede liberar bloqueos pero no obtener ninguno nuevo.

Inicialmente una transacción esta en fase de crecimiento, una vez que libere un bloqueo entra en fase de decrecimiento.

Este protocolo no evita interbloqueos ni retrocesos en cascada

b. Variantes

Protocolo de bloqueo estricto en dos fases

■ Los bloqueos exclusivos se liberan al final de la transacción

■ Se evitan los retrocesos es cascada.

Protocolo de bloqueo riguroso de dos fases

■ Todos los bloqueos se liberan al final de la transacción.

■ Las transacciones se pueden secuenciar en el orden que comprometen

12. Explique que es una Planificación Secuencial.

Consiste en una secuencia de instrucciones de varias transacciones, en la cual las instrucciones pertenecientes a una única transacción están juntas en dicha planificación.

De este modo, para un conjunto de n transacciones existen $n!$ planificaciones secuenciales válidas distintas.

13. Explique que es una Planificación Secuenciable.

Si una planificación P puede convertirse en una planificación secuencial P' , se dice que P es secuenciable.

14. Fundamente a su criterio la importancia de la Secuencialidad en SGBD.

Esta permite que no ocurran cuellos de botella, actualizaciones perdidas, lectura sucia entre otros que son un fastidioso y costoso a la hora de ponernos a solucionar. Siempre queremos que nuestro SGBD cumpla con los principios de atomicidad, consistencia, aislamiento y persistencia para con los datos. Es gracias a la secuencialidad podemos estar seguros que los cumple, pues va a gestionar ordenadamente las transacciones concurrentes evitando los problemas mencionados anteriormente.

15. Explique qué entiende por un gestor de bloqueos en un SGBD que implementa un Control de Concurrencia basado en Bloqueos. Indique además como él mismo podría estar implementado y que pasos debería de tener en cuenta para evitar la inanición de las transacciones.

Un control de concurrencia basado en bloqueos consiste en un mecanismo que permite reservar un elemento de la base de datos a una transacción dada y evitar que otras transacciones tengan acceso al mismo y pueden causar inconsistencias.

- Estos bloqueos están implementados para ejecutarse en dos modos: compartido y exclusivo. Los bloqueos compartidos permiten a un elemento bloqueado en una transacción T_1 ser leído por una transacción T_2 , pero no permite su modificación hasta que haya cesado el bloqueo. Los bloqueos exclusivos prohíben tanto la lectura como la escritura de un elemento bloqueado por una transacción a cualquier otra transacción que lo solicite.

- Se puede evitar la inanición de las transacciones al conceder los

bloqueos de la siguiente manera: cuando una transacción T_i solicita un bloqueo sobre un elemento de datos Q en un modo particular M , el gestor de control de concurrencia concede el bloqueo siempre que:

1. No exista otra transacción que posea un bloqueo sobre Q en un modo que esté en conflicto con M .
2. No exista otra transacción que esté esperando un bloqueo sobre Q y que lo haya solicitado antes que T_i .

16. **Al menos dos problemas o fenómenos de la concurrencia que deben ser evitados en Sistemas de BDs.**
17. **Un ejemplo de una planificación concurrente de dos transacciones no secuenciable en cuando a conflictos.**
18. **Un ejemplo de una planificación concurrente de dos transacciones secuenciable en cuando a vistas.**
19. **Esquematice el protocolo de compromiso de 2 fases indicando el nombre de cada fase. Asimismo, comente cómo se procede: (a) en caso de falla del coordinador y (b) en caso de falla de un sitio participante.**
20. **Explique cómo puede ser implementada una tabla de bloqueos en SGBDs con soporte para transacciones concurrentes.**
21. **Ilustre el diagrama de estados de una transacción explicando a qué corresponde cada estado y las diferentes transiciones posibles entre los mismos.**
22. Explique el Protocolo de Control de Concurrencia basado en Marcas de Tiempo que puede ser implementado en SGBDs.

NOSE

1. Ejemplifique la estrategia de la Semi Reunión para el procesamiento distribuido de reuniones. Explique al menos una ventaja de la misma que no se refiera a la transmisión de datos.

2. Teniendo en cuenta la consulta "select * from cuenta where saldo > 2500" y los siguientes datos estadísticos $B_r=100$, $N_r=2000$, $F_r=20$, $V(\text{saldo}, r)=20$. Estímese en forma explicada el costo medio y máximo en lectura de bloques de la consulta indicada. Debe asumirse que en la tabla existen valores de saldo menores, mayores e iguales a 2500.

Suponiendo que 2500 fuera el valor medio de los valores de la columna saldo significa que

hay 10 valores mayores a 2500. En promedio hay 100 filas con saldo mayor a 250, teniendo

en cuenta que hay 100 filas en promedio por cada valor del saldo. En promedio hay 50

bloques con filas cuyo saldo es mayor a 2500, ya que en promedio hay 20 filas por bloque.

Con el índice se puede planificar para que solo se realizan lecturas de estos 50 bloques. La

selección entonces toma como coste promedio la lectura de 50 bloques.

Realizar la proyección toma un tiempo de CPU igual a leer 1 bloque. El coste es igual a 51

bloques en promedio. El coste máximo es 101 bloques.

3. fdsaf

Tema 3 – 5 p. – Dadas las relaciones $r_1(A, B, C)$ y $r_2(C, D, E)$ con las siguientes propiedades: r_1 tiene 20.000 tuplas, r_2 tiene 60.000 tuplas, 100 tuplas de r_1 caben en un bloque y 30 tuplas de r_2 que caben en un bloque. Estímese el número de accesos a bloques requeridos utilizando las siguientes estrategias para calcular la operación de reunión

" $r_1 \text{ join } r_2 \text{ on } (r_1.C = r_2.C)$ ":

- a- Reunión en bucle anidado.
- b- Reunión en bucle anidado por bloques.
- c- Reunión por mezcla, suponiendo además que la relación r_1 debe ser ordenada externamente disponiéndose de 3 bloques de memoria intermedia.
- d- Reunión en bucle anidado indexada, suponiendo que existe un índice primario de árbol B+ en la columna C de la relación r_2 con nodos de 80 punteros.
- e- Reunión por Hash, suponiendo un tamaño de memoria principal disponible de $M = 15$.

4. fsdfsdasd

Tema 3 – 10 p. – Dadas las relaciones $r1(A, B, C)$ y $r2(C, D, E)$ con las siguientes propiedades: $r1$ tiene 40.000 tuplas, $r2$ tiene 90.000 tuplas, 100 tuplas de $r1$ caben en un bloque y 45 tuplas de $r2$ que caben en un bloque. Estímese el número de accesos a bloques requeridos utilizando las siguientes estrategias para la reunión " $r1 \text{ join } r2 \text{ on } (r1.C = r2.C)$ ":

- a- Reunión en bucle anidado.
- b- Reunión en bucle anidado por bloques.
- c- Reunión por mezcla, suponiendo además que la relación $r1$ debe ser ordenada externamente disponiéndose de 3 bloques de memoria intermedia.
- d- Reunión en bucle anidado indexada, suponiendo que existe un índice primario de árbol B+ en la columna C de la relación $r2$ con nodos de 60 punteros.

5. dfads

Dadas las relaciones $r1(A, B, C)$ y $r2(C, D, E)$ con las siguientes propiedades: $r1$ tiene 40.000 tuplas, $r2$ tiene 90.000 tuplas, 100 tuplas de $r1$ caben en un bloque y 45 tuplas de $r2$ que caben en un bloque. Estímese el número de accesos a bloques requeridos utilizando las siguientes estrategias para la reunión " $r1 \text{ join } r2 \text{ on } (r1.C = r2.C)$ ":

- a- Reunión en bucle anidado.
- b- Reunión en bucle anidado por bloques.
- c- Reunión por mezcla, suponiendo además que la relación $r1$ debe ser ordenada externamente disponiéndose de 3 bloques de memoria intermedia.
- d- Reunión en bucle anidado indexada, suponiendo que existe un índice primario de árbol B+ en la columna C de la relación $r2$ con nodos de 60 punteros.

$$nr1 = 40.000 \quad fr1 = 100 \quad br1 = 400$$

$$nr2 = 90.000 \quad fr2 = 45 \quad br2 = 2000$$

$$a. 400 + 40.000 * 2.000 = 80.000.400$$

$$b. 400 + 400 * 2000 = 800.400$$

$$c. 400 + 2.000 + 400 * (2 * \lceil \log_2 (400 / 3) \rceil + 1) = 9.200$$

$$d. 400 + 40.000 * (\lceil \log_{30} (90.000) \rceil + 1) = 200.400$$

6. dfasd

Tema 5– 10 p. – Dadas las relaciones $A(a1, a2, \dots, a20)$, $B(b1, b2, \dots, b12)$ y $C(c1, c2, \dots, c15)$, y la siguiente consulta:

```
select distinct A.a1, C.c1 from A join B {A.a2 = B.b3}
join C {C.c2 = B.b4} where A.a1 > 10 and B.b1 = 50,
```

Muestre:

1. Su traducción directa al álgebra relacional.
2. Los pasos para llegar a la expresión equivalente que minimice la cantidad de datos procesado por cada operación.
3. El árbol de evaluación de la expresión final.

7. fdasd

Tema 9 – 5p – Dadas las relaciones A(a1, a2, ..., a20), B(b1, b2, ..., b12) y C(c1, c2, ..., c15), y la siguiente consulta:

```
select distinct A.a1, C.c1 from A join B on (A.a2 = B.b3)
join C on (C.c2 = B.b4) where A.a1 > 10 and B.b1 = 50;
```

Muestre:

1. Su traducción directa al algebra relacional.
2. Detalle al menos 3 optimizaciones basadas en expresiones equivalente que pueden aplicarse. Debe indicar la regla de equivalencia aplicada.
3. El árbol de evaluación de la expresión final, conforme a su respuesta del punto anterior