

**Tema 1 – 20 p.** – Explique brevemente:

- a- En que consiste la estrategia RAID y detalle los Niveles 0, 1, 5, 1+0, 0+1.
- b- Los tipos de índices estudiados conforme su organización física y cuando conviene aplicarlos.
- c- Los pasos para el Procesamiento de Consultas. Haga el diagrama y explique cada paso.
- d- Los enfoques o estrategias de organización de registros en archivos de datos en un SGBD.
- e- Que características de los Árboles B y B+ los hace apropiados como estructura para implementar índices ordenados.

**Tema 2 – 5 p.** – 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   |

**Tema 3 – 5 p.** – Dadas las relaciones  $r1(A, B, C)$  y  $r2(C, D, E)$  con las siguientes propiedades:  $r1$  tiene 20.000 tuplas,  $r2$  tiene 60.000 tuplas, 100 tuplas de  $r1$  caben en un bloque y 30 tuplas de  $r2$  que caben en un bloque. Estímese el número de accesos a bloques requeridos utilizando las siguientes estrategias para calcular la operación de 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 80 punteros.
- e- Reunión por Hash, suponiendo un tamaño de memoria principal disponible de  $M = 15$ .

**Tema 4 – 5 p.** – 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:

- 1- Selección basada em Búsqueda Lineal.
- 2- Búsqueda Binaria.
- 3- Búsqueda en índice primario para un atributo clave.
- 4- Búsqueda en índice secundario para atributo no clave.

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

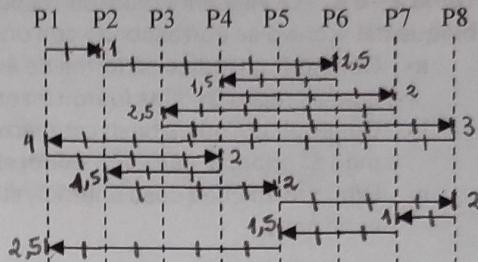
```
select distinct A.a1, C.c1 from A join B on (A.a2 = B.b3)
join C on(C.c2 = B.b4) where A.a1 > 10 and B.b1 = 50;
```

Muestre:

1. Su traducción directa al álgebra relacional.
2. Los pasos para optimizar la traducción original por medio de expresiones de equivalencias tal que minimice la cantidad de datos procesados en cada operación.
3. El árbol de evaluación de la expresión final.

**Tema 6 – 5 p.** – 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.



**Tema 7 – 5 p.** Dada la consulta `select * from A a join B 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.

UNIVERSIDAD NACIONAL DE ASUNCIÓN  
Facultad Politécnica

Bases de Datos II - Segundo Examen Parcial – 13/06/2024 – Duración 120 min.

**Tema 1 – 10 p. – Explique apropiadamente:**

- / a- Las formas de Almacenamiento Distribuido en Sistemas de BDDs, y sus ventajas.
- / b- Que se entiende por Alta Disponibilidad, Reconfiguración y Recuperación en Sistemas de BDDs.
- / c- El modelo de ejecución de transacciones distribuidas en Sistemas de BDDs.
- / d- Las propiedades de la transacciones que deben ser aseguradas por los SGBDs.
- / e- El protocolo del control de concurrencia en dos fases.

**Tema 2 – 5 p. – Protocolo de Quorum de Consenso. Explique o indique:**

- / a- El protocolo de control de bloqueo distribuido de Quórum de Consenso,
- / b- Que implica el que se cumpla cada una de las condiciones de quorum del protocolo, y
- / c- Las configuraciones de pesos que permiten emular los protocolos de Mayoría y Sesgado con el mismo.

**Tema 3 – 5 p. – Explique apropiadamente:**

- a- Al menos dos problemas o fenómenos de la concurrencia que deben ser evitados en Sistemas de BDs.
- / b- Un ejemplo de una planificación concurrente de dos transacciones no secuenciable en cuanto a conflictos.
- / c- Un ejemplo de una planificación concurrente de dos transacciones secuenciable en cuanto a vistas.

**Tema 4 – 5 p. – Estrategia de la Semireunión.** (a) Muestre los pasos para la reunión de las siguientes relaciones mediante la estrategia de la semireunión. Luego, (b) compare la misma en cuanto a la transmisión de información entre sitios respecto de la estrategia más simple que realiza el movimiento de relaciones entre los sitios. (c) Explique cuales son la principales ventajas que se consiguen con la aplicación de esta técnica.

| R@site1 |    |    | S@site2 |    |    |    |
|---------|----|----|---------|----|----|----|
| A1      | A2 | A3 | B1      | B2 | B3 | B4 |
| 1       | 6  | 2  | 5       | 1  | 9  | 6  |
| 2       | 7  | 4  | 6       | 2  | 0  | 5  |
| 2       | 8  | 6  | 7       | 3  | 1  | 4  |
| 4       | 7  | 5  | 9       | 4  | 2  | 3  |
| 7       | 5  | 3  | 1       | 9  | 6  | 3  |

R <join> S on (R.A2 = S.B1) @ site1

**Tema 5 – 5 p. – Ilustre y describa la arquitectura de implementación de un Dataware.**

**Tema 6 – 5 p. – Explique las 3 V's que caracterizan a las iniciativas de Big Data.**

**Tema 7 – 5 p. – Esquematice el protocolo de compromiso de 2 fases indicando el nombre de cada fase. Asimismo, comente como se procede: (a) en caso de falla del coordinador y (b) en caso de falla de un sitio participante.**

**Tema 8 – 5 p. – Explique cómo puede ser implementada una tabla de bloqueos en SGBDs con soporte para transacciones concurrentes.**

**Tema 9 – 5 p. – Ilustre el diagrama de estados de una transacción explicando a que corresponde cada estado y las diferentes transiciones posibles entre los mismos.**

UNIVERSIDAD NACIONAL DE ASUNCIÓN  
FACULTAD POLITÉCNICA  
INGENIERÍA EN INFORMÁTICA

BASES DE DATOS II – 1er. Examen Final – 04/07/2024 – Duración 120 min. – 18.00hs

Tema 1 – 5p – Explique las formas de Almacenamiento Distribuido que pueden ser implementadas por SGBDDs.

Replicación  
Fragmentación

Tema 2 – 5p – Explique el Protocolo de Control de Conurrencia basado en Marcas de Tiempo que puede ser implementados en SGBDs.

Tema 3 – 5p – Explique las cuatro (4) formas de organización de los registros de datos en los de Archivos de Datos.

Head  
sequential  
Hash  
B+ tree

Tema 4 – 5p – Detalle la estructura e implementación de registros de datos de longitud variable de estructura de páginas con ranuras.

Tema 5 – 5p – 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 estímese:

- El número de pistas procesadas asumiendo que no se aplica el algoritmo del ascensor.
- El número de pistas procesadas asumiendo la ampliación del algoritmo del ascensor.

Tema 6 – 5p – Explique la arquitectura de implementación de un Data Warehouse.

ETL DW Integración y operaciones H. de consultas e informes

H. EIS . TI - DATA

Tema 7 – 5p – Explique las 3 V's que caracterizan a las iniciativas de Big Data.

Volumen Variedad Velocidad

Tema 8 – 5p – Explique los (3) tres tipos de índices estudiados de acuerdo con su implementación física. Indique además para cada uno para qué tipo de consultas resultan más apropiados.

ordenados asociativos bitmap

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;

$\theta_1$        $\theta_2$

Muestre:

1. Su traducción directa al álgebra relacional.
2. Detalle al menos 3 optimizaciones basadas en expresiones equivalentes 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

AB ja BC

Tema 10 – 5p – 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.

UNIVERSIDAD NACIONAL DE ASUNCIÓN  
FACULTAD POLÍTÉCNICA  
INGENIERÍA EN INFORMÁTICA

BASES DE DATOS II – 1er. Examen Final – 04/07/2024 – Duración 120 min. – 18.00hs

Tema 1 – 5p – Explique las formas de Almacenamiento Distribuido que pueden ser implementadas por SGBDDs.

Replicación.  
Fragmentación.

Tema 2 – 5p – Explique el Protocolo de Control de Concurrencia basado en Marcas de Tiempo que puede ser implementados en SGBDs.

Tema 3 – 5p – Explique las cuatro (4) formas de organización de los registros de datos en los de Archivos de Datos.

Heap  
scattered  
Hash, un activo.  
obj. un activo.

Tema 4 – 5p – Detalle la estructura e implementación de registros de datos de longitud variable de estructura de páginas con ranuras.

Tema 5 – 5p – 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 estímese:

- 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.

Tema 6 – 5p – Explique la arquitectura de implementación de un Data Warehouse.

ETL DW Infraestructura y operaciones H. de consultas e informes. TI-EIS. TI-OLAP

Tema 7 – 5p – Explique las 3 V's que caracterizan a las iniciativas de Big Data.

Volumen Variedad Velocidad

Tema 8 – 5p – Explique los (3) tres tipos de índices estudiados de acuerdo con su implementación física. Indique además para cada uno para qué tipo de consultas resultan más apropiados.

ordenados asociativos bitmap

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 álgebra relacional.
2. Detalle al menos 3 optimizaciones basadas en expresiones equivalentes 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

20 7a 5b

Tema 10 – 5p – 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.

Temario 1: Explique:

a- Las medidas de rendimiento de los discos magnéticos.

• Tiempo de Acceso: Tiempo que toma leer o escribir datos desde su petición hasta su confirmación.

- Tiempo de búsqueda: Tiempo que se tarda el plato en ubicar la pista correcta. 4 a 10 ms.

- Latencia Rotacional: Tiempo que tarda en aparecer el sector a escribir/leer debajo de los cabezales. 4 a 15 ms.

• Tasa de Transferencia: Volumen de datos por unidad de tiempo que se pueden guardar o recuperar del disco.

• Tiempo medio de fallo (TMDF): Lapso de tiempo que un disco puede trabajar continuamente sin fallas.

b- Los enfoques de organización de archivos un SGBD.

• Heap: Los registros son guardados en cualquier lugar del archivo en donde existe espacio suficiente.

• secuencial: Los registros se guardan en orden secuencial de acuerdo a una clave de búsqueda.

• Hash: Los registros son almacenados en un bloque específico por una función hash que se aplica sobre el valor de un campo del registro.

• Organización en agrupación: los datos de diferentes relaciones se guardan en el mismo archivo.

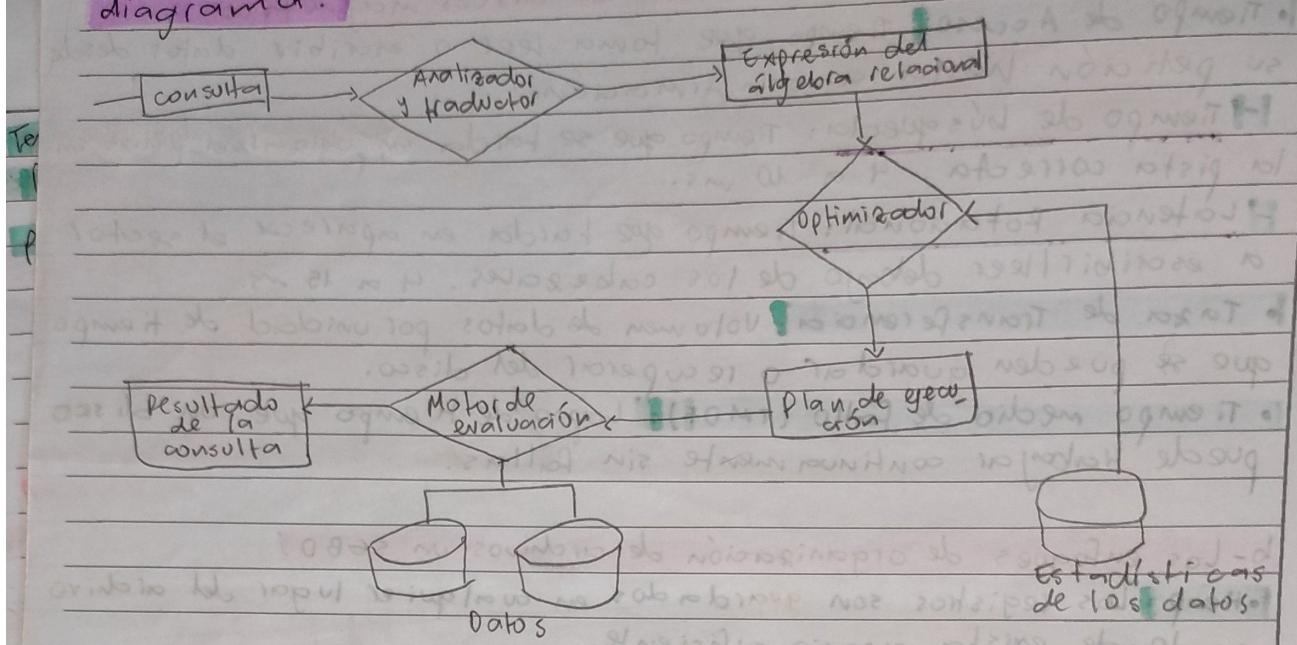
c- Los tres tipos de índices estudiados y cuando conviene aplicarlos.

• Índices Ordenados: Los registros índices están ordenados de acuerdo al valor de la clave de búsqueda. Existen dos tipos: Índices primarios e índices secundarios. Y conviene aplicarlos cuando en las consultas se utilizan un rango de valores.

• Índices Asociativos: Organizan las claves de búsqueda, con sus punteros asociados, dentro de una estructura de archivo asociativo (bucket). conviene aplicarlos cuando la consulta utiliza una condición de igualdad en la clave.

• Índices Bitmaps: Son un tipo especial de índice para consultar sobre múltiples claves de manera eficiente. Utilizan un array de bits, donde cada bit indica la presencia o ausencia de ciertas características en una colección de registros. Y conviene utilizarlos cuando se realizan consultas de selección con múltiples condiciones, especialmente en conjunción con claves de búsqueda.

d- Los pasos para el procesamiento de consultas - Haga el diagrama.



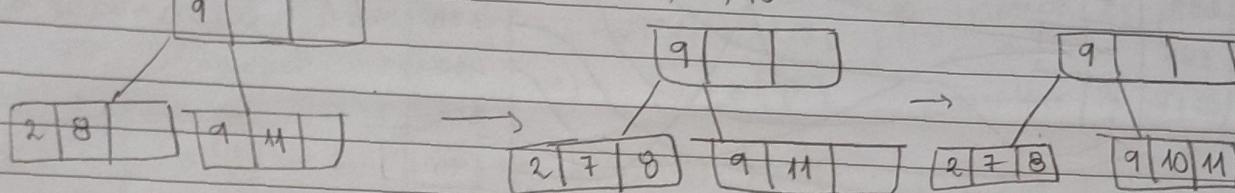
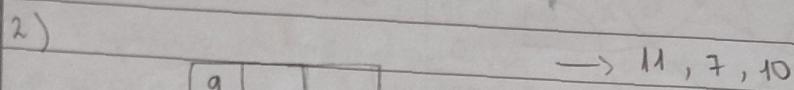
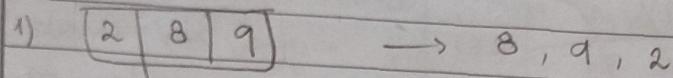
- Análisis y traducción: se revisa la sintaxis y se verifica en el catálogo del sistema que todos los identificadores sean nombres de objetos de la base de datos.
- Optimización: En esta fase se obtiene como resultado el plan de ejecución de la consulta, haciendo uso de la estadística de datos.
- Evaluación: Se evalúan las operaciones del plan de ejecución, ya sea con la Materialización (una operación a la vez), o con el Encuadramiento (varias operaciones simultáneamente).

Y se realiza la traducción a expresiones del álgebra relacional

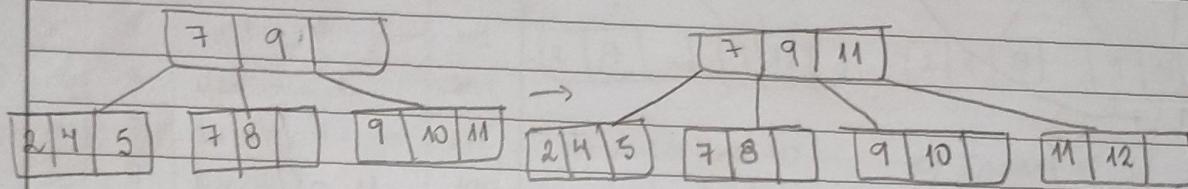
- Análisis y traducción: se revisa la sintaxis y se verifica en el catálogo del sistema que todos los identificadores sean nombres de objetos de la base de datos. Y se realiza la traducción a expresiones del álgebra relacional.
- Fase de optimización: se obtiene como resultado el plan de ejecución de la consulta, haciendo uso de la estadística de datos.
- Evaluación: Se evalúan las operaciones del plan de ejecución, ya sea con la Materialización (una operación a la vez), o con el Encuadramiento (varias operaciones simultáneamente).

Tema 2: La sgte. tabla corresponde al archivo de la relación diente, en el que cada bloque del archivo corresponde a 1 fila, se pide:

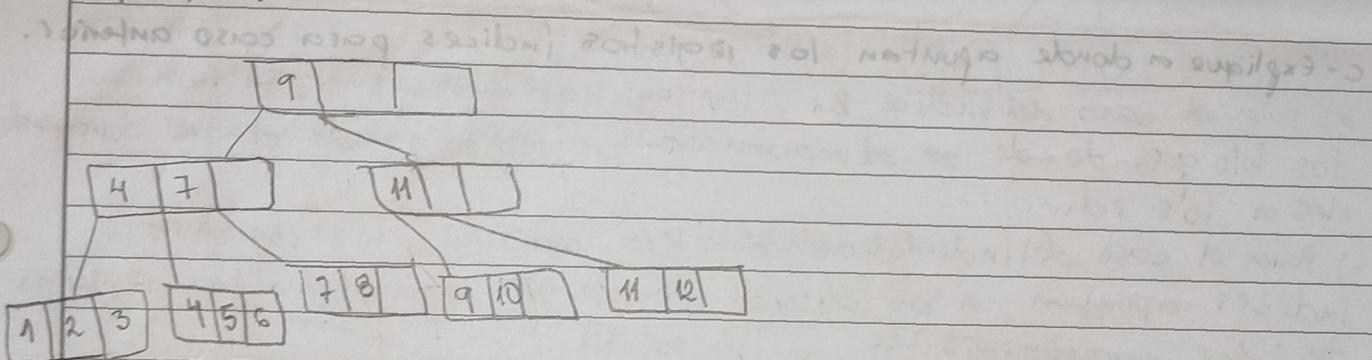
a- construir un índice en forma de árbol B+ con nodos de 4 punteros para la clave primaria id, suponiendo que los registros fueron insertados según el orden de los valores de la columna nombre.



3) → 4, 5, 12



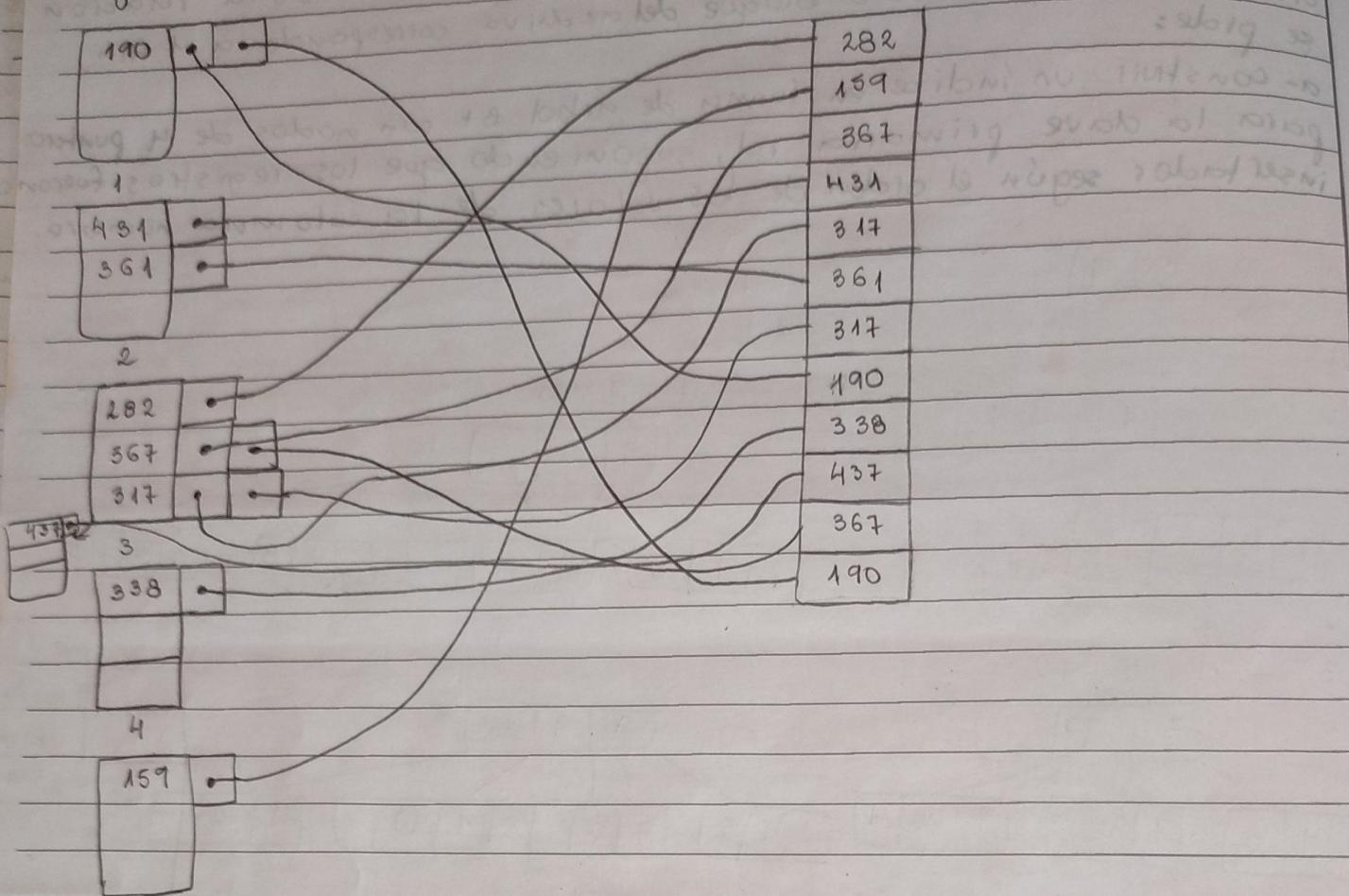
4) → 1, 3, 6



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.

$$a \bmod m = a - m \lfloor a/m \rfloor$$

$$a \bmod m = a - m \lfloor \frac{a}{m} \rfloor$$



c- Explique a donde apuntan los registros indices para caso anterior

a) Para el caso del indice B+, los registros indices apuntan a los bloques donde se almacenan los registros de datos correspondientes a los id.

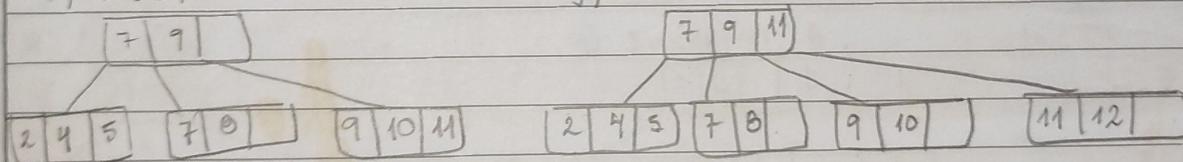
b) Para el caso del indice hash estático cerrado, los registros indices apuntan a los cajones donde se encuentran los registros de datos correspondientes a los saldos.

Tema 1: La sgte. tabla corresponde al estado actual del archivo de la relación cliente, en el que cada bloque del archivo corresponde a 1 fila, se pide:

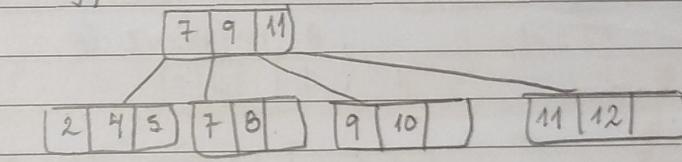
- Construir un índice conforme de árbol B+ con nodos de 4 punteros para la clave primaria ID, suponiendo que los registros/filas fueron insertados según el orden alfabético de la columna nombre.
- Construir un índice hash estático cerrado con ocho slots de 4 elementos cuya función de asociación es " $x \bmod 4$ " sobre la columna saldo, siendo  $x$  el valor de cada fila en dicha columna.
- Explique detalladamente en cada caso si el índice es primario o secundario. ABCDEFGHIJKLMNOPQRSTUVWXYZ.

| ID   | nombre      | saldo | a) | 2 8 9   | → Insertar 8, 9, 2 |
|------|-------------|-------|----|---------|--------------------|
| X 1  | Preston S.  | 282   | b) | 9       | Insertar 11        |
| X 2  | cathleen S. | 159   | c) | 9       | Insertar 7         |
| X 3  | Tatyana R.  | 367   |    | 2 8     |                    |
| X 4  | Libby M.    | 431   |    | 9 11    |                    |
| X 5  | Oida R.     | 317   | d) | 2 7 8   | Insertar 10        |
| X 6  | Vivian C.   | 361   | e) | 9 11    | Insertar 4         |
| + 7  | Kirk T.     | 317   |    | 9       |                    |
| X 8  | Amanda M.   | 190   | f) | 7 9     | Insertar 5         |
| X 9  | Barry M.    | 398   |    | 1       |                    |
| X 10 | Lee L.      | 437   | g) | 2 7 8   | Insertar 12        |
| X 11 | Elliot F.   | 367   |    | 9 10 11 |                    |
| X 12 | Paula J.    | 190   |    | 2 4     |                    |
|      |             |       |    | 7 8     |                    |
|      |             |       |    | 9 10 11 |                    |

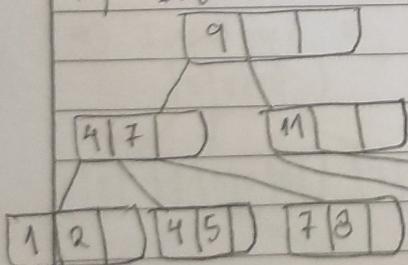
f) Insertar 5



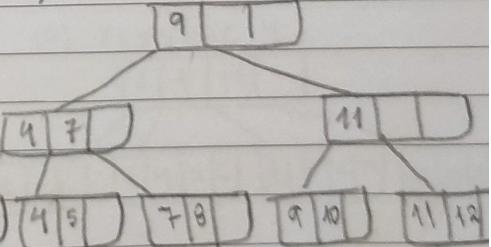
g) Insertar 12



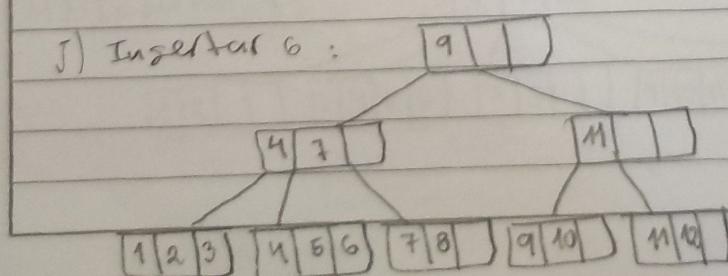
h) Insertar 1



i) Insertar 3

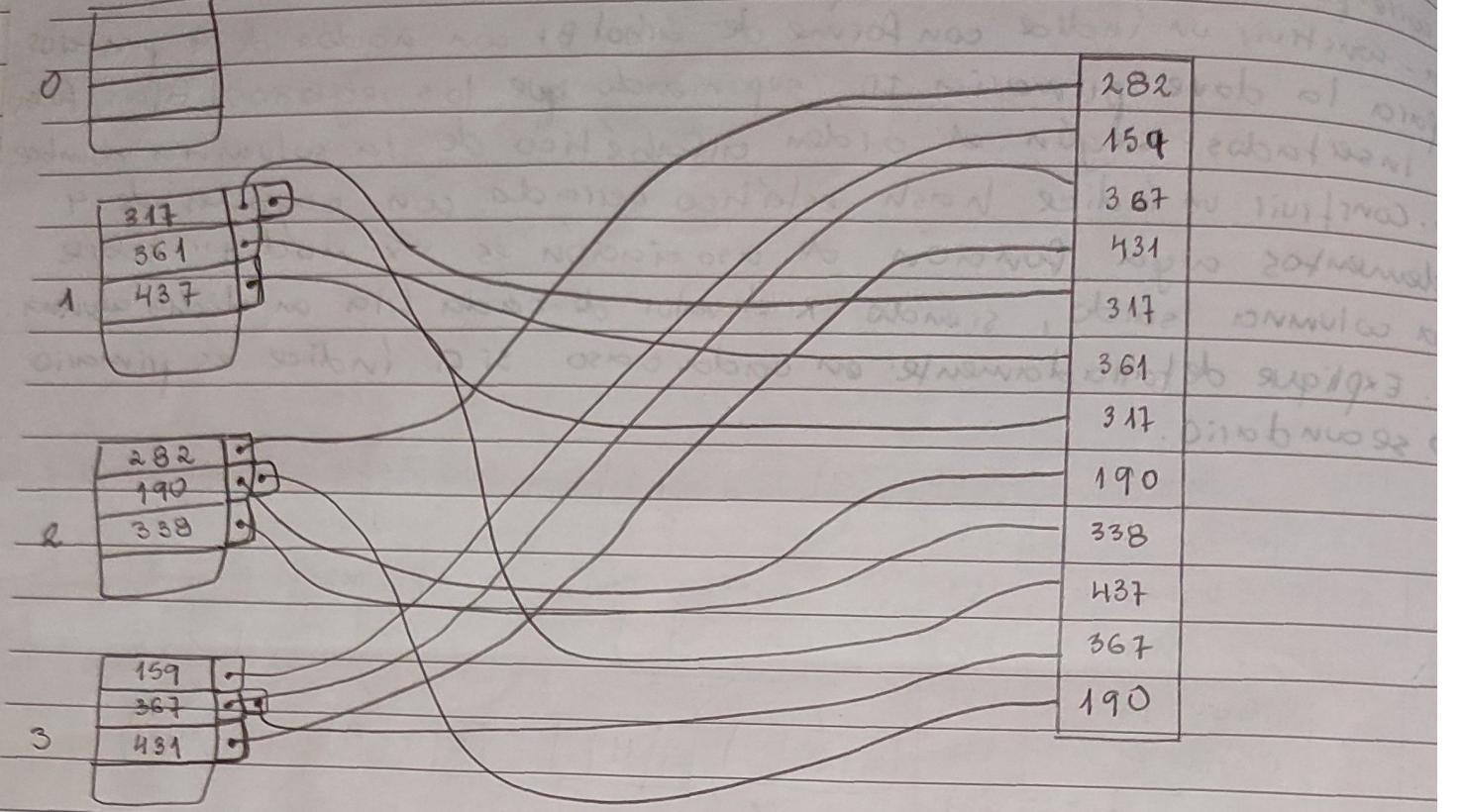


j) Insertar 6 :



6) " $x \bmod 4$ "  $\rightarrow a \bmod m = a - m \left\lfloor \frac{a}{m} \right\rfloor$

cajones de 4 elementos.



c) En el caso del árbol B+, el índice es primario porque el mismo ordena la tabla de manera secuencial.

• En el caso del hash estático cerrado, el índice es secundario porque el mismo no ordena la tabla de manera secuencial.

Tema 10: a) construir árbol BT según el orden de los valores de la columna nombre. (8 puntos)

b) construir un índice hash estático cerrado con cajones de 5 elementos cuya función de evaluación es " $x \bmod 4$ " sobre la columna saldo cuando  $x$  el valor de el fila en dicha columna.

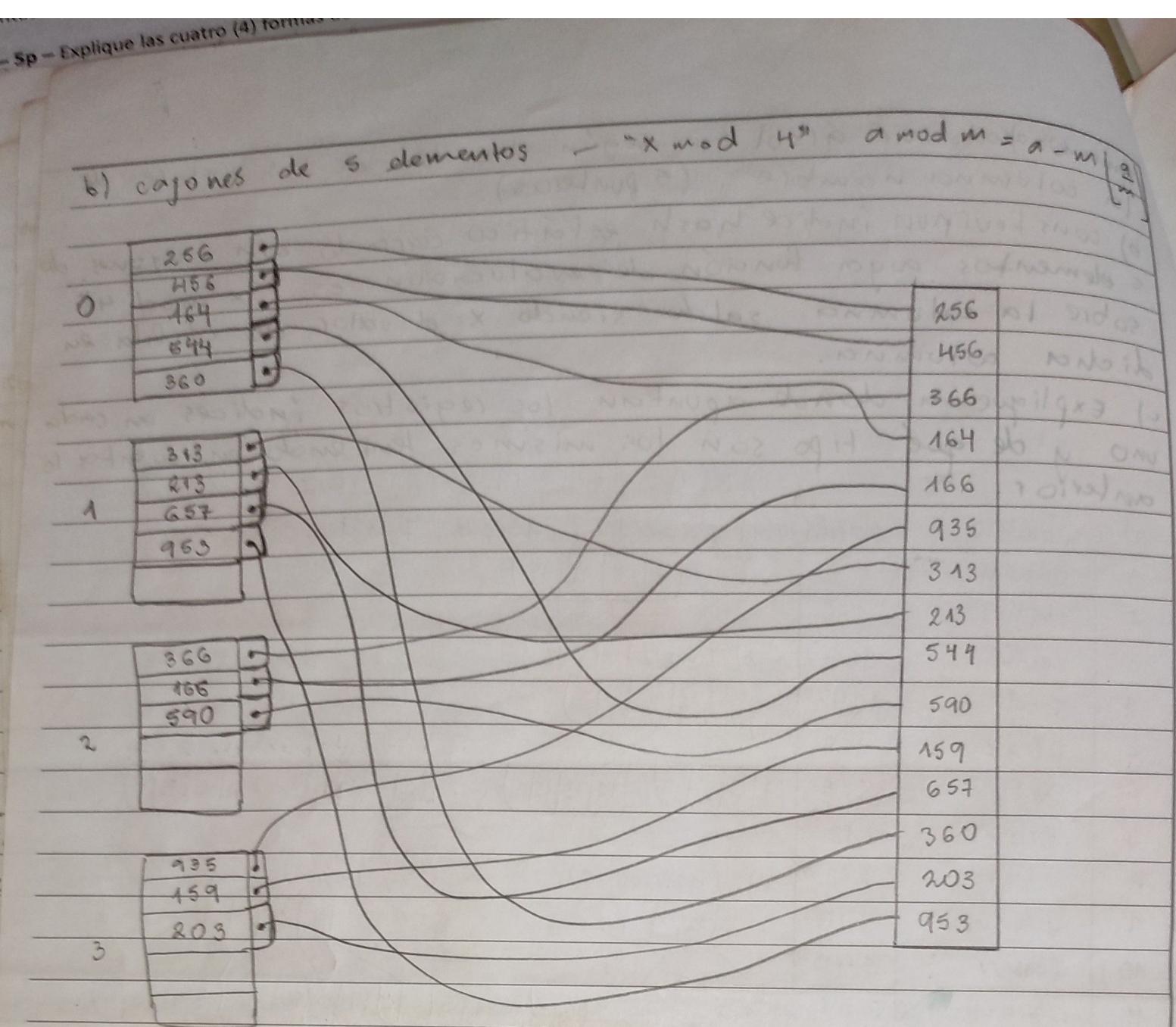
c) Explique a donde apuntan los registros índices en cada uno y de qué tipo son los mismos teniendo en cuenta lo anterior.

ABCD EFGHIJKLMNOPQRSTUVWXYZ

| id | nombre   | saldo |  |
|----|----------|-------|--|
| 1  | griffin  | 286   |  |
| 2  | Phillip  | 456   |  |
| 3  | Kevin    | 366   |  |
| 4  | Tad      | 164   |  |
| 5  | Derek    | 166   |  |
| 6  | Option   | 935   |  |
| 7  | Broni    | 313   |  |
| 8  | Axel     | 213   |  |
| 9  | Wayne    | 544   |  |
| 10 | Foller   | 590   |  |
| 11 | Mehill   | 159   |  |
| 12 | Chadwick | 657   |  |
| 13 | David    | 360   |  |
| 14 | Takeen   | 203   |  |
| 15 | Andrew   | 953   |  |

Diagrama de inserciones:

- 1) Insertar 15, 8, 7, 12
- 2) Insertar 13
- 3) Insertar 5
- 4) Insertar 10
- 5) Insertar 1
- 6) Insertar 14
- 7) Insertar 3
- 8) Insertar 11
- 9) Insertar 6
- 10) Insertar 4
- 11) Insertar 9;



c) a) Para el caso del índice B+, los registros índices apuntan a los bloques donde se almacenan los registros de datos correspondientes a los id. El índice es primario porque ordena la tabla de manera secuencial.

b) Para el caso del índice hash estático cerrado, los registros índices apuntan a los cajones donde se encuentran los registros de datos correspondientes a los saldos. El índice es secundario porque el mismo no ordena la tabla de manera secuencial.

Tema 3: Dadas las relaciones  $r_1(A, B, C)$  y  $r_2(C, D, E)$  con las sgtes. propiedades:  $r_1$  tiene 40.000 tuplas,  $r_2$  tiene 90.000 tuplas, 100 tuplas de  $r_1$  caben en un bloque y 45 tuplas de  $r_2$  que caben en un bloque. Estírese el número de accesos a bloques requeridos utilizando las sgtes estrategias para la reunión  $r_1 \text{ join } r_2 \text{ on } (r_1.C = r_2.C)$

- Reunión en bucle anidado.
- Reunión en bucle anidado por bloques.
- Reunión por mezcla, suponiendo además que la relación  $r_1$  debe ser ordenada externamente disponiéndose de 3 bloques de memoria intermedia.
- 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 60 punteros.

$$n_{r1} = 40.000$$

$$f_{r1} = 100$$

$$b_{r1} = \left\lceil \frac{n_{r1}}{f_{r1}} \right\rceil \rightarrow b_{r1} = \left\lceil \frac{40.000}{100} \right\rceil = 400$$

$$n_{r2} = 90.000$$

$$f_{r2} = 45$$

$$b_{r2} = \left\lceil \frac{n_{r2}}{f_{r2}} \right\rceil \rightarrow b_{r2} = \left\lceil \frac{90.000}{45} \right\rceil = 2000$$

a) Reunión en bucle anidado:

$$\begin{aligned} \text{costo} &= n_{r1} * b_{r2} + b_{r1} \\ &= 40.000 * 2000 + 400 \\ &= 80.000.400 \end{aligned}$$

b) Reunión en bucle anidado por bloques:

$$\begin{aligned} \text{costo} &= b_{r1} * b_{r2} + b_{r1} \\ &= 400 * 2000 + 400 \\ &= 800.400 \end{aligned}$$

c) Reunión por mezcla

$$\begin{aligned} \text{costo} &= b_{r1} + b_{r2} + \text{costo ordenación} \\ &= 400 + 2000 + 6800 \\ &= 9200 \end{aligned}$$

d) Reunión en bucle anidado indexada

$$\begin{aligned} \text{costo} &= b_{r1} + n_{r1} * (\lceil \log_2(\frac{n_{r2}}{b_{r2}}) \rceil + 1) \\ &= 400 + 40.000 * (\lceil \log_2(\frac{90.000}{60}) \rceil + 1) \\ &= 400 + 40.000 * 5 \end{aligned}$$

$$\begin{aligned} \text{costo ordenación} &= b_{r2} * (2 * \lceil \log_{M-1}(b_{r1}/M) \rceil + 1) \\ &= 600 * (2 * \lceil \log_2(400/3) \rceil + 1) \\ &= 600 * (2 * 8 + 1) \\ &= 6800 \end{aligned}$$

$$= 200.400$$

$n_r \rightarrow$  N° de tuplas de  $r$ .  
 $f_r \rightarrow$  N° de tuplas que caben en un bloque  
 $b_r \rightarrow$  N° de bloques que contienen tuplas de  $r$   
 $\left\lceil \frac{n_r}{f_r} \right\rceil$

• Reunión por hash:

• Si  $N \geq M$ :

$$(b_r + b_s) * (2 * \lceil \log_{M-1}(b_s) - 1 \rceil + 1)$$

• Si  $N < M$ :

$$3(b_r + b_s) + 2N$$

$$N = \frac{b_r}{M}$$

Tema 4) Dada una operación de selección cualquiera basada en una condición de igualdad, explique en qué casos un sebo utilizar los sgtes algoritmos e indique cuál es el costo asintótico de los mismos.

1- Búsqueda Lineal

2- Búsqueda Binaria

3- Búsqueda en índice primario para un atributo clave.

4- Búsqueda en índice secundario para atributo no clave.

1- Búsqueda Lineal: se puede aplicar a cualquier archivo sin importar la ordenación del mismo o la presencia de índices, pero es más lento que otros algoritmos de búsqueda, ya que lee todos los bloques del archivo.

$$\text{costo} = b_1$$

2- Búsqueda Binaria: se puede aplicar si el archivo está ordenado según un atributo y la condición de igualdad implica dicho atributo.

$$\text{costo} = \lceil \log_2(b_1) \rceil + \lceil c_s(A, r) / f_r \rceil - 1$$

3- Búsqueda en índice primario para un atributo clave:

- se puede utilizar cuando el índice está construido sobre el atributo clave, por lo que se necesita leer un solo bloque.

$$\text{costo} = AA_i + 1$$

4- Búsqueda en índice secundario para atributo no clave:

- se puede utilizar si el campo indexado no es una clave, por lo que se explora el índice y se accede a un cajón de  $c_s(A, r)$  punteros, en el peor de los casos cada uno de los registros se encuentra en diferentes bloques.

$$\text{costo} = AA_i + 1 + CS(A, r)$$

Tema 4: Dada la consulta de abajo, proporcione una traducción inicial de la misma en álgebra 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 e.EMPLEADO = e.ID
    join PROYECTO p on (p.ID = te.PROYECTO)
    where p.NOMBRE = 'AQUARIUS' and e.FECHA_NAC >= '2000-01-01'
  
```

a) Caso 1

Traducción:  $\Pi_{e.LNAME} (\sigma_{\theta_1 \wedge \theta_2} (p \bowtie_{p.ID=te.PROYECTO} (e \bowtie_{te.EMPLEADO=e.ID})))$

se aplica:

- Regla 7.b: se distribuye la operación de selección de forma separada a  $\theta_1$  que implica únicamente a  $p$  y a  $\theta_2$  que implica únicamente a  $e$ .

$$\Pi_{e.LNAME} (\sigma_{\theta_1} (p) \bowtie_{p.ID=te.PROYECTO} \sigma_{\theta_2} (e \bowtie_{te.EMPLEADO=e.ID}))$$

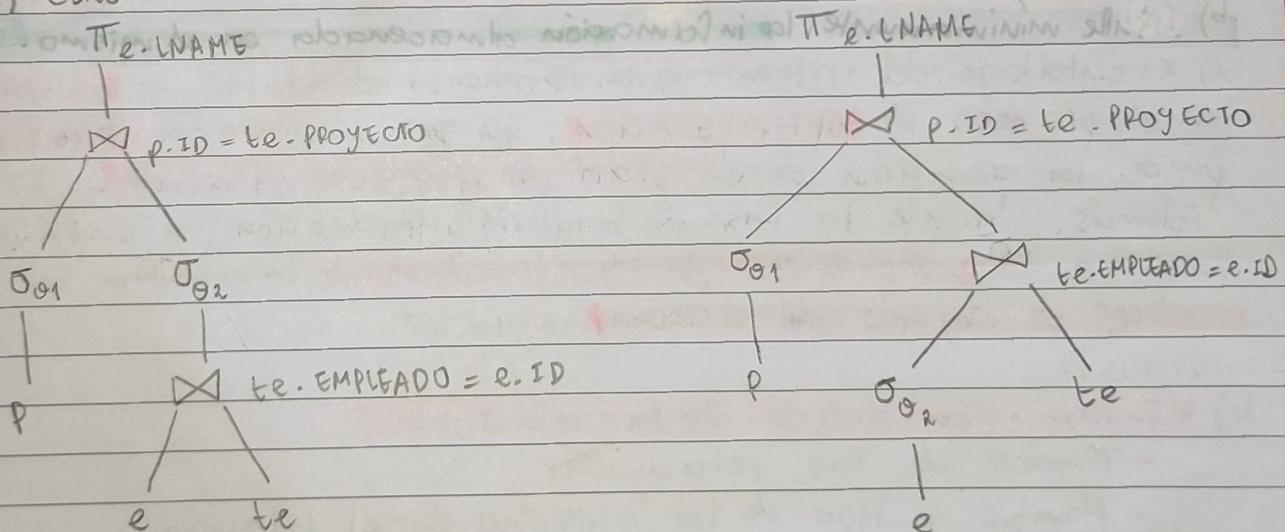
b) Caso 2, se puede seguir optimizando el caso 1.

se aplica:

- Regla 7.a: se distribuye la operación de selección sobre  $\theta_2$  que solo involucra a  $e$ .

$$\Pi_{e.LNAME} (\sigma_{\theta_1} (p) \bowtie_{p.ID=te.PROYECTO} (\sigma_{\theta_2} (e) \bowtie_{te}))$$

c) Caso 1



6.b: son asociaitivos ( ).

7: La operación de selección se distribuye sobre  $\bowtie$  cuando:

- la condición de selección solo involucra ~~atributos de~~ una de las relaciones (en este caso E1).

$$\sigma_{\theta_0} (E_1 \bowtie E_2) = (\sigma_{\theta_0} (E_1)) \bowtie E_2$$

- cuando las condiciones de selección involucran de forma separada a atributos de un ~~a~~ de relaciones.

$$\sigma_{\theta_1 \wedge \theta_2} (E_1 \bowtie E_2) = (\sigma_{\theta_1} (E_1)) \bowtie (\sigma_{\theta_2} (E_2))$$

Tema 9) Explique el porqué la organización basada en árboles B y árboles B+ resulta apropiada como estructuras de datos para implementar índices ordenados.

- La organización basada en árboles B y árboles B+ proporciona una estructura de datos eficiente y equilibrada para la implementación de índices ordenados, lo que facilita la búsqueda y recuperación de datos.

• Explique las medidas de rendimiento que deben tener en cuenta para la elección de unidades de discos magnéticos. ¿A su criterio, cuál de las mismas sería la más determinante?

- Tiempo de Acceso:
  - Tiempo de búsqueda: - - -
  - Latencia Rotacional: - - -
- Tasa de Transferencias: - - -
- Tiempo medio de fallos: - - -

A mi criterio, el tiempo medio de fallos sería la más determinante, porque indica la confiabilidad y durabilidad a largo plazo del dispositivo.

a) Indique la importancia del catálogo del sistema (Diccionario de datos).

b) Detalle mínimamente la información almacenada en él mismo.

a) El catálogo del sistema o diccionario de datos es crucial en el proceso de optimización, ya que proporciona datos estadísticos para la creación de un plan de ejecución eficiente. Además, durante la fase de análisis y traducción, el catálogo facilita la comparación de identificadores de la consulta con los nombres de objetos del sistema.

b) • Información acerca de las relaciones:

- Nombre de las relaciones.
- Nombre y tipo de los atributos de la relación.
- Nombre y definiciones de las vistas.
- Restricciones de integridad.

• Cuentas de usuario, incluyendo nombre y contraseña.

• Información estadística y descriptiva:

- N° de tuplas de la R.
- Método de almacenamiento y manejo de cada relación.

• Información de la organización física:

- Tipo de almacenamiento de la relación.
- Localización física de cada relación.

• Información de los índices.

- Nombre del índice y de su relación.
- Tipo de índice
- Atributos sobre los cuales se crea el índice.

Toma 9) Explique el porqué la organización basada en árboles B y árboles B+ resulta apropiada como estructura de datos para implementar índices ordenados.

- La organización basada en árboles B y árboles B+ proporciona una estructura de datos eficiente y equilibrada para la implementación de índices ordenados, lo que facilita la búsqueda y recuperación de datos.

Explique las medidas de rendimiento que deben tener en cuenta para la elección de unidades de discos magnéticos. ¿A su criterio, cuál de las mismas sería la más determinante?

- Tiempo de Acceso:
  - Tiempo de búsqueda: - - -
  - Latencia Rotacional: - - -
  - Tasa de transferencias: - - -
  - Tiempo medio de fallos: - - -

A mi criterio, el tiempo medio de fallos sería la más determinante, porque indica la fiabilidad y durabilidad a largo plazo del dispositivo.

a) Indique la importancia del catálogo del sistema (Diccionario de datos).

b) Detalle mínimamente la información almacenada en él mismo.

a) El catálogo del sistema o diccionario de datos es crucial en el proceso de optimización, ya que proporciona datos estadísticos para la creación de un plan de ejecución eficiente. Además, durante la fase de análisis y traducción, el catálogo facilita la comparación de identificadores de la consulta con los nombres de objetos del sistema.

- b)
- Información acerca de las relaciones:
    - Nombre de las relaciones.
    - Nombre y tipo de los atributos de la relación.
    - Nombre y definiciones de las vistas.
    - Restricciones de integridad.

• Cuentas de usuario, incluyendo nombre y contraseña.

|   |   |  |
|---|---|--|
| • Información estadística y descriptiva:  | • Información de la organización física:  | • Información de los índices.  |
| <ul style="list-style-type: none"> <li>- N° de tuplas de la R.</li> <li>- Método de almacenamiento de cada relación.</li> </ul> | <ul style="list-style-type: none"> <li>- Tipo de almacenamiento de la relación.</li> <li>- Localización física de cada relación.</li> </ul> | <ul style="list-style-type: none"> <li>- Nombre del índice y de sus relaciones.</li> <li>- Tipo de índice.</li> <li>- Atributos sobre los cuales se crea el índice.</li> </ul> |

Tema 5) Dadas las relaciones  $A(a_1, a_2, \dots, a_{20})$ ,  $B(b_1, b_2, \dots, b_{12})$  y  $C(c_1, c_2, \dots, c_{15})$ , 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. Sustitución directa al álgebra relacional.
2. Los pasos para llegar a la expresión equivalente que minimice la cantidad de datos procesados por cada operación.
3. El árbol de evaluación de la expresión final.

$$\Pi_{A.a1, C.c1} (\sigma_{\theta_1} \wedge \sigma_{\theta_2} (A \bowtie_{A.a2=B.b3} (B \bowtie_{C.c2=B.b4} C)))$$

2. se aplica:

- Regla 7.b: se distribuye la operación de selección de forma separada a  $\theta_1$  que implica únicamente a A y a  $\theta_2$  que implica únicamente a B.

$$\Pi_{A.a1, C.c1} (\sigma_{\theta_1}^{(A)} \bowtie_{A.a2=B.b3} \sigma_{\theta_2}^{(B)} \bowtie_{C.c2=B.b4} C)$$

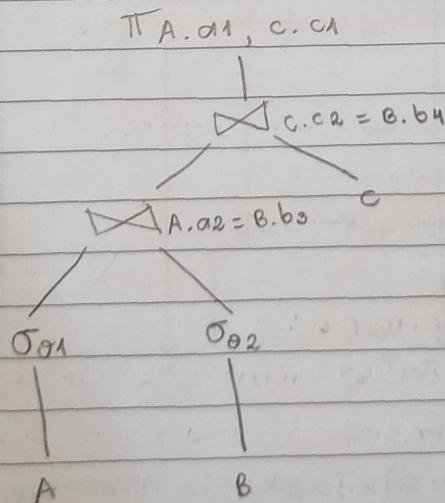
- Regla 7.a: se distribuye la operación de selección sobre  $\theta_2$  que solo involucra a B.

$$\Pi_{A.a1, C.c1} (\sigma_{\theta_1}^{(A)} \bowtie_{A.a2=B.b3} (\sigma_{\theta_2}^{(B)} \bowtie_{C.c2=B.b4} C))$$

- Regla 6.a: La operación de  $\bowtie$  es asociativa.

$$\Pi_{A.a1, C.c1} ((\sigma_{\theta_1}^{(A)} \bowtie_{A.a2=B.b3} \sigma_{\theta_2}^{(B)}) \bowtie_{C.c2=B.b4} C)$$

3.



Tema 8- La sgte. figura que muestra la planificación de lectura del disco en un SED que no utiliza el algoritmo del ascensor. se pide:

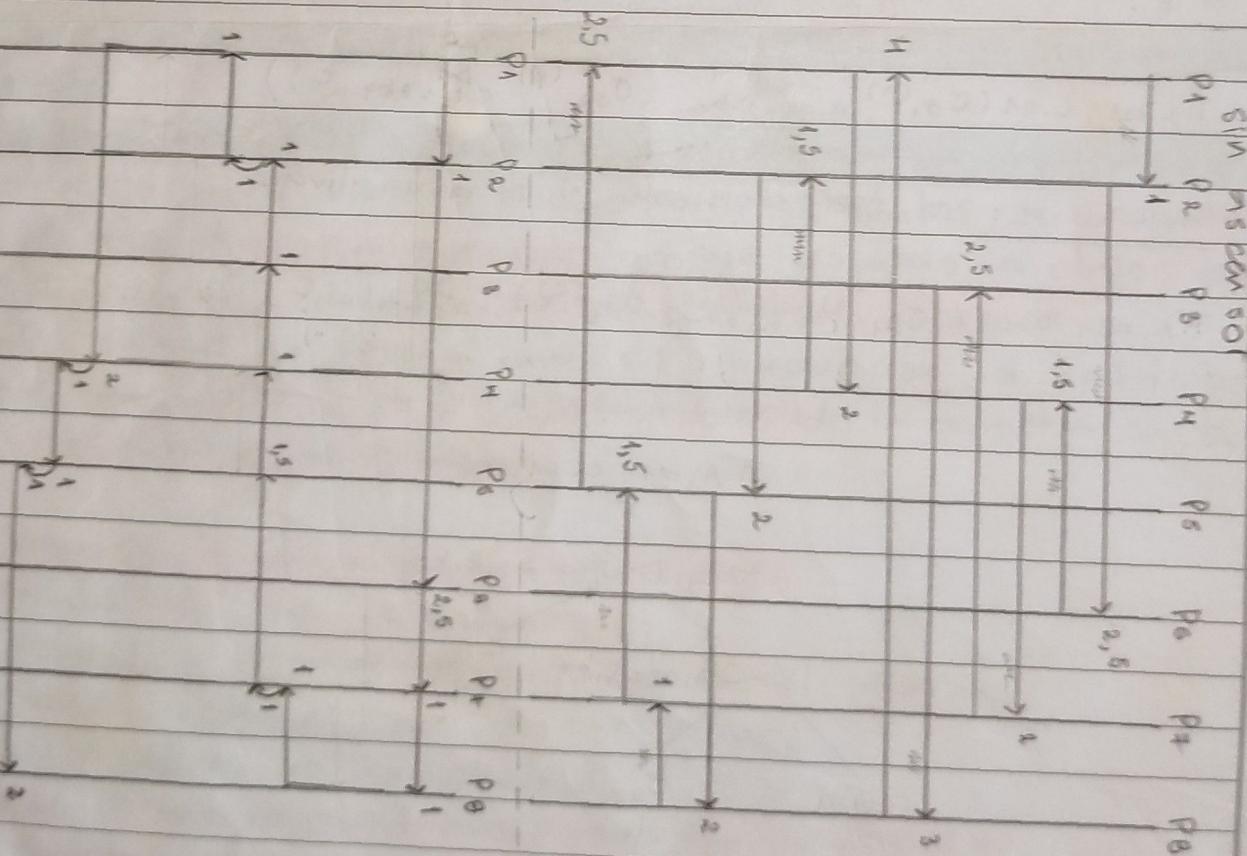
- a- Explique el algoritmo del ascensor y el porqué de su utilidad.  
b- Muestre una planificación basada en el uso del algoritmo del ascensor suponiendo que:

1. El recorrido del peine empieza en la pista 1.
2. Se pueden atender varias solicitudes por paradas.
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.

a- El algoritmo del ascensor funciona moviéndose de manera secuencial hacia adelante o hacia atrás según las solicitudes de acceso, atendiendo primero las más cercanas y evitando movimientos innecesarios de ida y vuelta. Esto reduce el tiempo de espera y mejora el rendimiento del sistema de almacenamiento.

b)



Tiempo total sin ascensor = 29 ms      c) Taza sin ascensor =  $\frac{29}{14} = 2,07$  solicitudes/ms

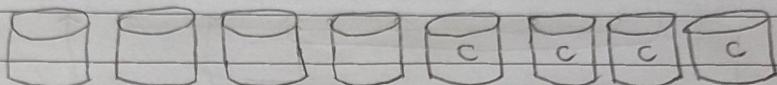
Tiempo total con ascensor = 21 ms

Taza con ascensor =  $\frac{21}{14} = 1,5$  solicitudes/ms

TEMA 2: Explique detalladamente las ventajas de configurar un espacio de almacenamiento basado en RAID. Ejemplifique además los niveles 1 y 5. (Redundant Array of Independent Disks)

- Esta configuración ofrece la posibilidad de mejorar la fiabilidad del almacenamiento de datos, ya que se puede guardar información repetida en varios discos, por lo que el fallo de un disco no provoca pérdida de datos.

RAID de nivel 1: se refiere a la creación de imágenes del disco con distribución de bloques. Es ideal para casos en donde la seguridad de los datos es más importante que el desempeño de las escrituras.

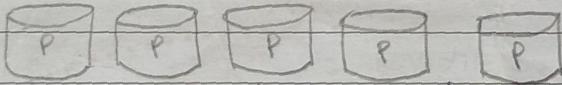


- Discos con imagen

RAID de nivel 5: Paridad distribuida con bloques entrelazados

Divide los datos y la paridad entre los  $N+1$  discos, en vez de guardar los datos en  $N$  discos y la paridad en uno.

Todos los discos pueden participar en la atención a las solicitudes de lectura, por lo que aumenta el número total de solicitudes que pueden atenderse en una cantidad de tiempo dada.



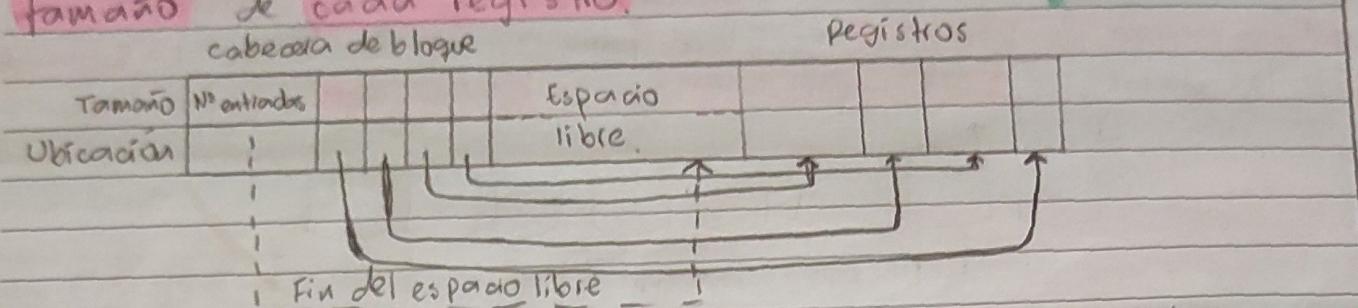
- Paridad distribuida con bloques entrelazados.

RAID 0: No proporciona redundancia, por lo que si un disco falla, se pierden todos los datos.

Tema 9) Detalle apropiadamente como se implementa la organización física de bloques denominada estructura de páginas por ranuras para el almacenamiento de registros.

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

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



- Los registros reales se ubican en el bloque de manera contigua empezando por el final del bloque.
- El espacio libre 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.
- Si se borra un registro, se libera el espacio que ocupa y se da el valor de borrado a su entrada (el valor de -1 a su tamaño). Además, se desplazan los registros del bloque situados antes del registro borrado. Se actualiza el puntero de final del espacio libre en la cabecera.
- La estructura de páginas con ranuras necesita que no haya punteros que apunten directamente a los registros.

1) Explique:

2) Las formas de Almacenamiento Distribuido en BDD y sus ventajas.

- **Réplica de datos:** si la relación  $R$  se replica, se guarda una copia de dicha relación en dos o más sitios. Existe el caso de una réplica completa, en la que se guarda una copia en cada sitio del sistema.

- Ventajas:

- **Disponibilidad:** la falla del sitio donde se ubica la relación  $R$  no resulta en la indisponibilidad de los datos de  $R$ .
- **Paralelismo:** Múltiples consultas de lectura sobre  $R$  pueden ser procesadas en paralelo por varios sitios.
- **Transferencia de Datos Peculiares:** solo las actualizaciones en  $R$  provocan que varios datos sean transferidos por la red.

- Desventajas:

- **Costo de Actualización Incrementado:** cada réplica de  $R$  debe ser actualizada, lo cual puede provocar un alto tráfico en la red.
- **Complejidad de Control de Conurrencia Aumentada:** La actualización concurrente de las réplicas requiere implementar mecanismos especiales de control de concurrencia.
- **Solución:** Utilizar copias primarias.
- **Fragmentación de Datos:** División de la relación  $R$  en fragmentos  $R_1, R_2, \dots, R_n$ , los cuales contienen información suficiente para重构  $R$ . Hay dos esquemas diferentes de fragmentación de las relaciones: fragmentación horizontal y fragmentación vertical. La fragmentación horizontal divide la relación descomponiendo el esquema  $R$  de la relación  $r$ .

- Ventajas:

• **Horizontal:**

- Permite el procesamiento en paralelo de los fragmentos.
- Permite que las tuplas se ubiquen el sitio donde son más frecuentemente accedidas.

• **Vertical:**

- Permite que los fragmentos sean almacenados en el sitio más apropiado para la información.
- Permite el procesamiento paralelo sobre una relación.

## Transparencia de Datos.

- Es el grado en el cual un usuario no tiene noción o no es afectado<sup>3)</sup> por los detalles en como fueron distribuidos y donde están almacenados los datos en un sistema distribuido.
- La Transparencia de Datos puede referirse a:
  - Transparencia de Fragmentación.
  - Transparencia de Réplica.
  - Transparencia de Ubicación.

## b) Que se entiende por Alta Disponibilidad y Reconfiguración en un sistema de BDD.

- **Alta Disponibilidad:** La mayor parte del tiempo el sistema debería estar disponible (ej: 99,99%).
- **Reconfiguración:** Re adaptar el sistema para continuar con el cómputo.
  - Debe basarse en:
    - Abortar las transacciones que estaban activas en el sitio que ha fallado.
    - Si el sitio que falla posee replicas de datos, el catálogo del sistema debe ser modificado para que no enviar transacciones a ese sitio hasta que se recupere.
    - Si ha fallado un sitio que hospedaba un servidor central de algún subsistema, debe elegirse un nuevo sitio que lo reemplace.

## c) El algoritmo del Luchador.

- Cuando un sitio si envía un mensaje al coordinador y este no responde en un intervalo de tiempo  $T$ , se asume que el coordinador ha fallado y él trata de convertirse en el nuevo coordinador.
- Si envía un mensaje de elección a todos los sitios con mayor identificación y espera un tiempo  $T$  por una respuesta.
  - cuando no hay una respuesta, si asume que todos los sitios de coordinador han fallado y se elige a sí mismo como coordinador.
  - cuando hay una respuesta, si aguarda un nuevo tiempo  $T'$  para recibir la respuesta del nuevo sitio con mayor identificación que ha sido electo.
  - cuando no se recibe un mensaje del nuevo coordinador en un tiempo  $T'$ , entonces si reinicia el algoritmo.
- Despues de que se haya recuperado un sitio, comienza de inmediato la ejecución de este mismo algoritmo.
  - si no hay ningún sitio activo con un número más elevado que el sitio recuperado, este obliga a todos los sitios menores a permitir transformarse en el nuevo sitio coordinador, aun que ya hay un coordinador activo con un número más bajo.

### Exerc 3) Explique:

- a) El protocolo de control de bloques distribuido de Quórum de consenso:
- El protocolo de consenso de quórum es una generalización del protocolo de Mayoría y del protocolo de sesgado. El protocolo de consenso de quórum asigna a cada sitio un peso no negativo. Asigna a las operaciones de lectura y de escritura del elemento  $x$  dos enteros, denominados quórum de lectura  $Q_l$  y quórum de escritura  $Q_e$ , que deben cumplir la condición siguiente, donde  $s$  es el peso total de todos los sitios en los que reside  $x$ :

$$Q_l + Q_e > s \quad y \quad 2 * Q_e > s$$

- b) Que implica el que se cumpla la condición del protocolo y:
- Que se cumpla la condición implica que cuando haya un bloqueo de lectura no se pueda realizar otro bloqueo de escritura, y viceversa; es decir, que los bloqueos compartidos y exclusivos sean incompatibles.

- c) como pueden emularse los protocolos de Mayoría y sesgado con el mismo:

Asumiendo que todos los pesos son iguales, para emular el protocolo de Mayoría de hace  $Q_l = Q_e = \frac{n}{2} + 1$ , para asegurar que más de la mitad de las réplicas estén bloqueadas.

Para emular el protocolo de sesgado se hace  $Q_l = 1$  y  $Q_e = n$ , ya que para la lectura se necesita solo un sitio, pero para la escritura se necesitan todos.

- 4) Esquematice el protocolo de compromiso de 2 fases e indique como se procede: (C2F).

a) En caso de falla del coordinador y.

b) en caso de un sitio participante.

Fase 1:

- Esta fase comienza luego de la última sentencia de T.
- ci añade el mensaje <Preparar T> a la bitácora y fuerza a guardar la bitácora en almacenamiento persistente.
- Ci envía el mensaje <Preparar T> a todos los participantes SK:
- En cada sitio participante SK:
  - El gestor GTK determina si se puede comprometer su parte de T.
  - Si no es posible comprometer T en SK:
    - GTK añade <No T> a su bitácora.
    - GTK responde a ci enviando el mensaje <Abortar T>

- Si es posible comprometer T en SK:
  - GTR añade «Preparado T» a su bitácora y fuerza su grabación.
  - GTR responde a C enviando el mensaje «Preparado T».

### Fase 2.

- Esta fase comienza:
  - cuando C recibe todas las respuestas «preparado T» de los SK's,
  - o cuando ha transcurrido cierto tiempo desde que se ha enviado el mensaje «preparar T» y no se han recibido todas las respuestas.
- Si se han recibido todas las respuestas de los SK's y todos son el mensaje «Preparado T»:
  - C añade el registro «comprometido T» a la bitácora y fuerza su grabación.
  - C envía el mensaje «comrometer T» a todos los SK's.
  - Cuando un SK recibe el mensaje «Comrometer T»
    - Compromete su parte de la transacción T.
    - Añade el registro «comprometido T» a su bitácora y fuerza su grabación.
- Si no se han recibido todas las respuestas de los SK's o una de ellas es «Abortar T»:
  - C añade el registro «Abortado T» a la bitácora y fuerza su grabación.
  - C envía el mensaje «Abortar T» a todos los SK's.
  - Cuando un SK recibe el mensaje «Abortar T»
    - Aborta su parte de la transacción T y ejecuta la recuperación.
    - Añade el registro «Abortado T» a su bitácora y fuerza su grabación.

- a) **Fallo del coordinador:** Si el coordinador falla durante la ejecución del protocolo de confirmación de T, los sitios participantes deben decidir el destino de T:
- Si un sitio SK contiene el registro «Comrometer T», entonces T debe ser comprometida.
  - Si un sitio SK contiene el registro «Abortar T», entonces T debe ser abortada.
  - Si ningún sitio contiene el registro «Preparado T», entonces C no ha decidido que T debe confirmarse. De esta forma T se aborta.
  - Si no ocurren los casos anteriores, entonces todos o algunos de los sitios tienen en su bitácora «Preparado T» y están esperando por la orden del coordinador. Los sitios deben esperar a que C se recupere para conocer el destino de T.

file transformarse en ~

un coordinador activo con un número m-

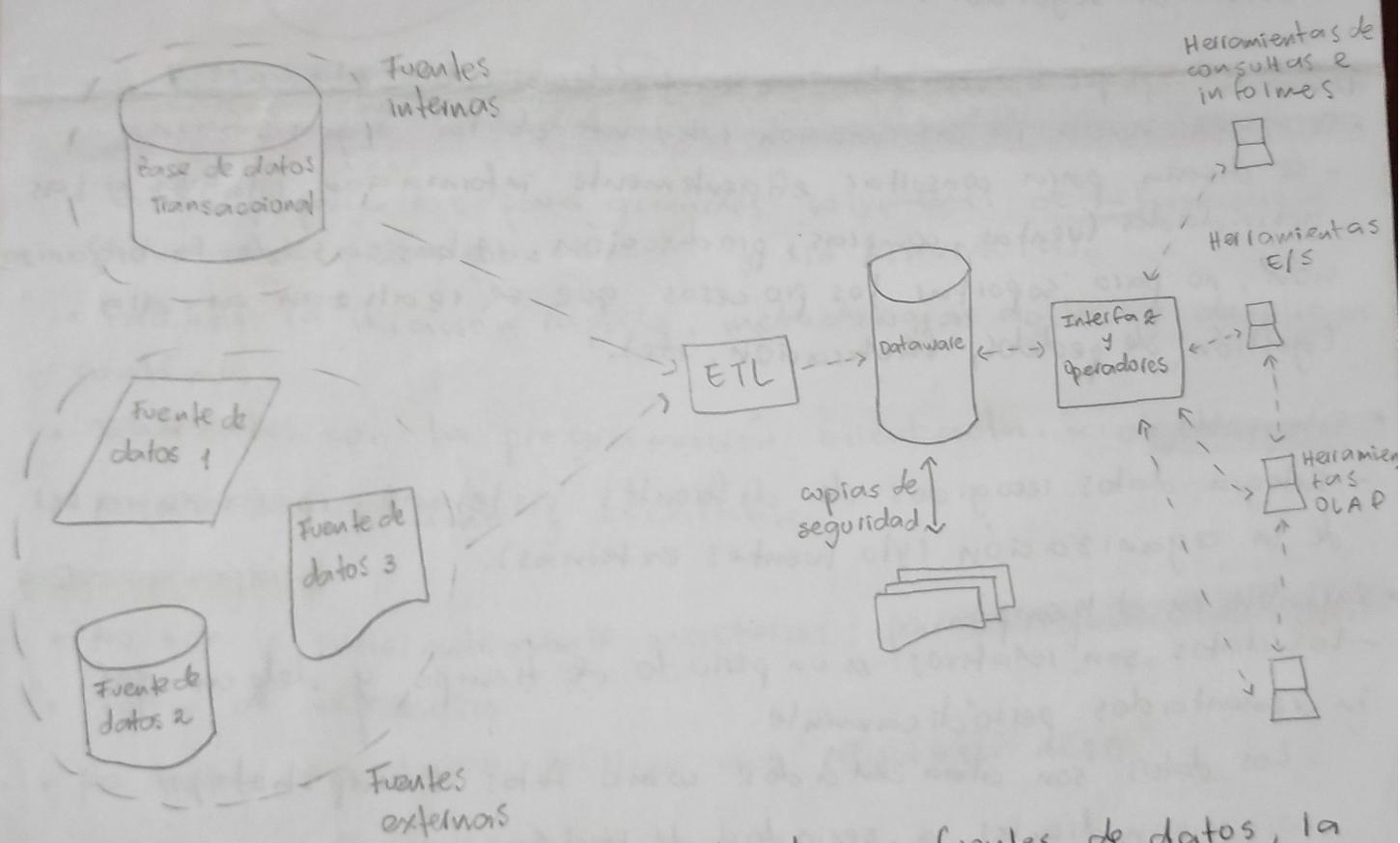
### Fallo de un sitio.

Si un sitio se cae antes, durante o después del protocolo C2F, al recuperarse este utiliza su bitácora para determinar el destino de T:

Ti

- Si contiene <comprobado T>, SR realiza rehacer (T).
- Si contiene <Abortar T>, SR realiza deshacer (T)
- Si contiene <Preparado T>, se consulta con el para determinar el destino de T.
  - Si el responde <comprobado T>, SR realiza rehacer (T).
  - Si el responde <Abortado T>, SR realiza deshacer (T).
- Si su bitácora no contiene ningún registro de T.
  - La falla de SR ocurre antes del protocolo de compromiso de T, por lo tanto a ha abortado T.
  - SR realiza deshacer (T).

### 5) Ilustre y describa la arquitectura de implementación de un Dataware.



- **ETL:** se encarga de la extracción de las fuentes de datos, la transformación y la carga de dataware.
- **Dataware:** Es la base de datos que se utiliza para la toma de decisiones y es constituida a partir de los registros de las transacciones del negocio.
- **OLAP:** Es una herramienta que realiza el análisis en línea de manera rápida de las informaciones multidimensionales.

6) Explique las 3 V's que caracterizan a los iniciativas de Big Data, IoT, IOT, d...

- **Volumen:**

- Grandes volúmenes de información.
- se está pasando de hablar en gigabytes o terabytes a tamaños de datos de Petabytes, Exabytes o Zettabytes. Volúmenes que nos escapan.

- **Variedad:**

- Información de tipos muy diversos.
- Ya no solo tenemos información estructurada en Bases de Datos o Archivos.
- Ahora empezamos a tener información con tipos diferentes y totalmente desestructurada.

- **Velocidad:**

- Velocidad con la que se genera esta información hace imposible gestionarla con sistemas de base de datos convencionales. Las empresas y las personas ya no quieren estar al día, quieren "estar al segundo".

7) Cite y explique brevemente 4 características del Dataware.

- **Orientado hacia la información relevante de la organización:**

- se diseña para consultar eficientemente información relativa a las actividades (ventas, compras, producción,...) básicas de la organización, no para soportar los procesos que se realizan en ella (gestión de pedidos, facturación, etc).

- **Integrado:**

- Integra datos recogidos de diferentes sistemas operacionales de la organización (y/o fuentes externas).

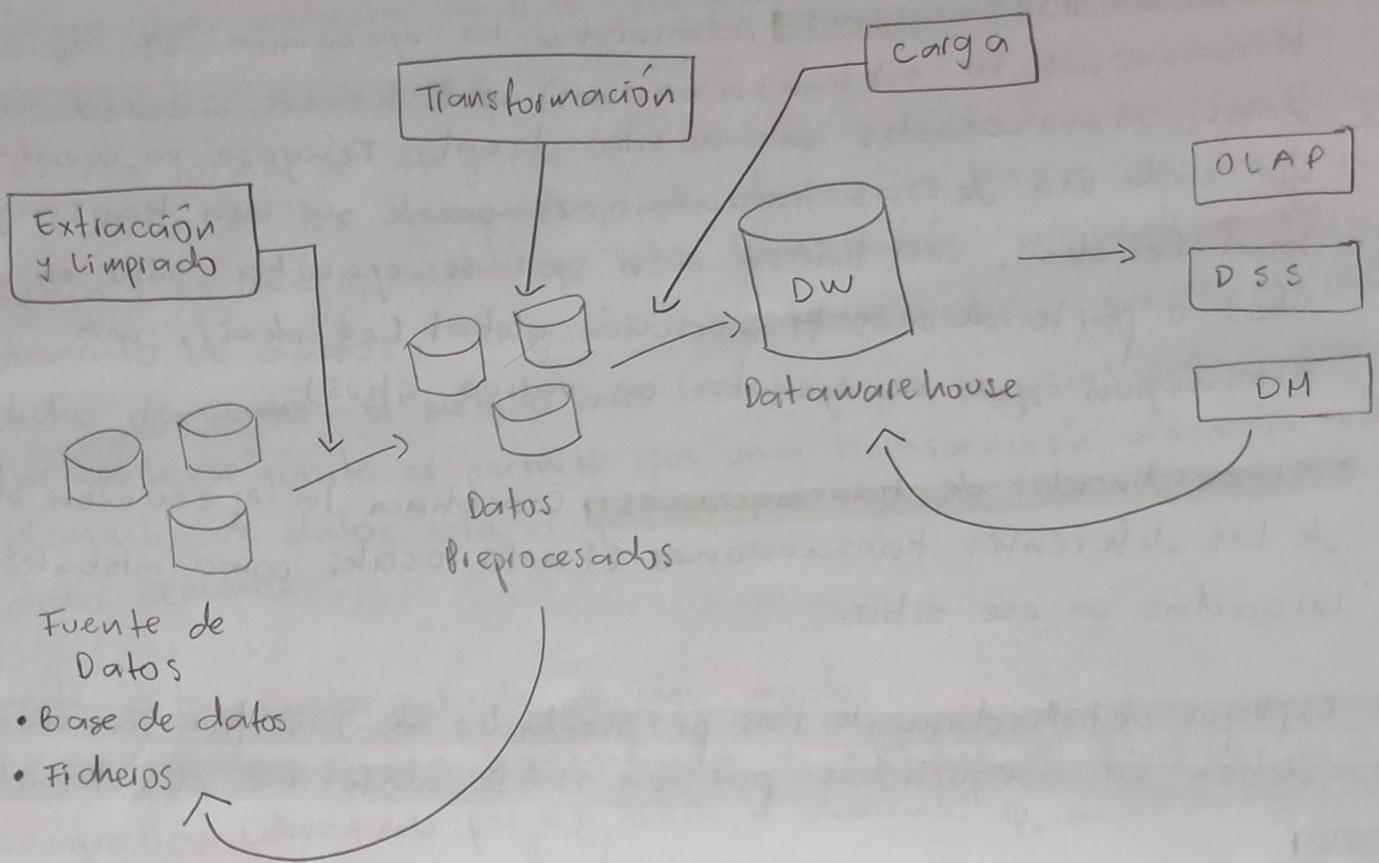
- **Variable en el tiempo:**

- Los datos son relativos a un periodo de tiempo y deben ser incrementados periódicamente.
  - Los datos son almacenados como fotos (snapshots) correspondientes a períodos de tiempo.

- **No volátil:**

- Los datos almacenados no son actualizados, sólo son incrementados.

Ilustre el proceso de extracción, transformación y carga (ETL) de datos en DW.



9) Cite las ventajas y las desventajas de una BD - No SQL.

- Ventajas:

- Están preparados para grandes volúmenes de información estructurada, semi estructurada y sin estructuras.
- Facilita la iteración rápida, metodologías ágiles y despliegue continuo.
- Coherentes con la programación orientada a objetos.
- Arquitectura eficiente y escalable.

- Desventajas:

- No son lo suficientemente maduras para algunas empresas.
- Falta de experiencia.
- No trata con datos críticos que requieren ACID.
- Problemas de compatibilidad.

10) Explique detalladamente el Modelo Abstracto del sistema de Transacciones en un sistema de Bases de Datos Distribuidas orientados a asegurar las propiedades ACID.

- **Gestor de transacciones:** Administra la ejecución de las transacciones (o subtransacciones) que tienen acceso a los datos almacenados en un sitio local. Téngase en cuenta que cada una de esas transacciones puede ser una transacción local (es decir, una transacción que se ejecuta sólo en ese sitio) o parte de una transacción global (es decir, una transacción que se ejecuta en varios sitios).

- **El coordinador de transacciones:** Coordina la ejecución de las diferentes transacciones (tanto locales como globales) iniciadas en ese sitio.

11) Explique detalladamente las propiedades de las transacciones que deben ser aseguradas por un SBD conforme el estándar SQL.

- Para preservar la integridad de los datos las sgtes. propiedades de las transacciones deben ser aseguradas por el SBD:

- **Atomicidad (Atomicity):** Una transacción es una unidad atómica de procesamiento; es realizada enteramente o no es realizada en nada.

- **Consistencia (Consistency):** La ejecución de la transacción debe pasar la BD desde un estado consistente a otro también consistente.

- **Aislamiento (Isolation):** Una transacción no deberá hacer visible sus modificaciones a otras transacciones hasta que esté confirmada.

- **Persistencia (Durability):** Cuando una transacción cambia la BD y los cambios son confirmados, estos cambios no deben perderse por fallos posteriores.

2) Explique qué entiende por un Gestor de Bloqueos en un SGBD que implementa un control de concurrencia basado en bloques. Indique además como el mismo podría estar implementando y que pasos debería de tener en cuenta para evitar la inanición de las transacciones.

- Es una forma de asegurar la secuencialidad, exigiendo que el acceso a los elementos de datos se haga en exclusión mutua; es decir, mientras una transacción accede a un elemento de datos, ninguna otra transacción puede modificar dicho elemento. El método más habitual que se usa para implementar este requisito es permitir que una transacción acceda a un elemento de datos sólo si posee actualmente un bloqueo sobre dicho elemento.
- Puede estar implementado de dos modos:
  1. Compartido: si una transacción  $T_i$  obtiene un bloqueo en modo compartido (denotado por  $c$ ) sobre el elemento  $Q$ , entonces  $T_i$  puede leer  $Q$  pero no lo puede escribir.
  2. Exclusivo: si una transacción  $T_i$  obtiene un bloqueo en modo exclusivo (denotado por  $x$ ) sobre el elemento  $Q$ , entonces  $T_i$  puede tanto leer como escribir  $Q$ .
- 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  $T_i$ . De este modo, una petición de bloqueo nunca se quedará bloqueada por otra petición de bloqueo solicitada más tarde.

13) Explique en el contexto de los protocolos del control de concurrencia que se entiende por:

(1) Planificación, (2) Planificación secuenciable, (3) Secuencialidad en cuanto a conflicto y (4) Secuencialidad en cuanto a vistas.

1. Planificación: consiste en un ordenamiento de todas las operaciones de las transacciones, pero conservando el orden interno en cada transacción individual.

2. Planificación secuenciable: la ejecución concurrente de varias transacciones es correcta si y sólo si su efecto es el mismo que el obtenido al realizar las mismas transacciones en forma secuencial. Un plan secuencial siempre es correcto.

- Un plan  $s$  de  $n$  transacciones es secuenciable si es equivalente a algún plan secuencial de las mismas  $n$  transacciones.
- Decir que un plan no secuencial  $s$  es secuenciable es equivalente a decir que  $s$  es correcto.

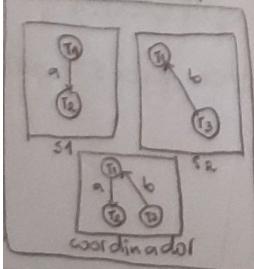
3. Secuencialidad en cuanto a conflictos: se dice que una planificación  $P$  es secuenciable en cuanto a conflictos, si es equivalente en cuanto a conflictos a una planificación secuencial.

4. Secuencialidad en cuanto a vistas: se dice que la planificación  $P$  es secuenciable en cuanto a vistas si es equivalente en cuanto a vistas a una planificación secuencial.

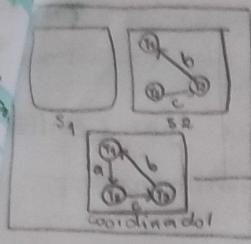
14) Explique en qué consisten los problemas que deben evitarse en la detección de interbloqueos en un sistema de Base de Datos Distribuida.

- En el enfoque de detección centralizada de interbloqueos el sistema crea y mantiene un grafo global de espera (la unión de todos los grafos locales) en un solo sitio: el coordinador de detección de interbloqueos. Si el coordinador halla un ciclo, selecciona una víctima para hacerla retroceder. Este esquema puede producir retrocesos innecesarios, en los casos de problemas de: ciclos falsos y retrocesos no relacionados.

#### Ciclos Falsos



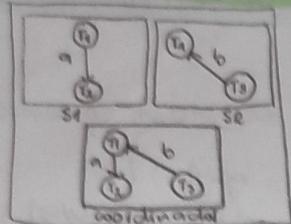
- Luego ocurre en el siguiente orden que:
1. T<sub>2</sub> libera el dato a
  2. S<sub>1</sub> envía el mensaje de eliminar el arco T<sub>1</sub> → T<sub>2</sub>
  3. T<sub>2</sub> solicita un bloqueo sobre c en el sitio S<sub>2</sub>, pero T<sub>3</sub> tiene bloq
  4. S<sub>2</sub> envía el mensaje de añadir el arco T<sub>2</sub> → T<sub>3</sub>.



se detecta un falso ciclo si:

1. El mensaje de s1 llega luego del mensaje de s2.
2. El mensaje de s1 se pierde.

### • Proceso No Relacionado:



2. S2 envía el mensaje de arriba al arco  $T_2 \rightarrow T_3$ .

Estado inicial

Luego ocurre en el segt. orden que:

1.  $T_2$  solicita un bloqueo sobre c en el sitio s2, pero  $T_3$  tiene bloqueado el dato c.

(5) Ilustre y describa la arquitectura de una BD Big Data.

Comente además acerca de 3 conceptos equivocados de una BD Big Data mencionados en clase.

bases de datos operacionales



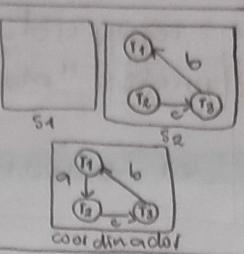
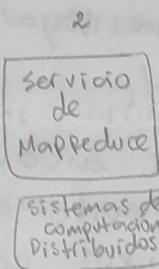
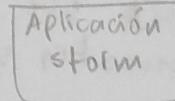
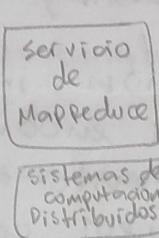
Datos de Internet

Archivos

Flujo

Multimedia

sensores

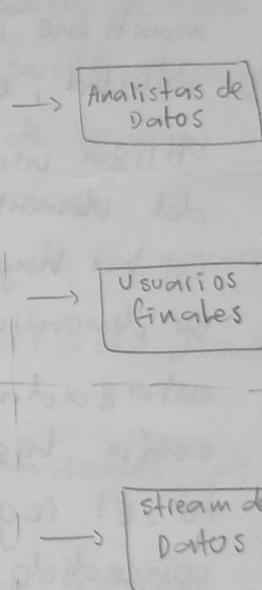


se detecta un ciclo falso si:

1. En s1 se aborta  $T_1$  por algún otro problema, no relacionado a la detección de ciclos.

Ocurre que:

1. El coordinador pudo haber elegido como víctima a  $T_2$  o a  $T_3$ .
2. se retrocede innecesariamente a  $T_2$  o a  $T_3$ .



Fuentes

Exploración

cloud ETL

Almacenamiento

- 1) Procesamiento en la nube.
- 2) Procesos Map Reduce.
- 3) Carga.

### - ¿Qué no es Big Data?

- No es una base de datos enorme.
- No es un datawarehouse enorme.
- No es una nueva forma de Business Intelligence.
- No es llevar las bases de datos a la nube.

DA  
Explique las formas de Almac  
16) Una de las características de las BD NoSQL, corresponde a) Indique la escalabilidad Horizontal. Describir el concepto de la misma trans

- Escalabilidad Horizontal: La implementación típica se realiza en muchos nodos de capacidad de procesado limitada, en vez de utilizar grandes "mainframes".

17) Describa la función del componente de gestión de concurrencia en un SEBD.

- Se encarga de asegurar que el efecto de la ejecución concurrente de un conjunto de transacciones sea igual a que las transacciones se hubiesen ejecutado secuencialmente (lo cual siempre es correcto).

18) Describa la estructura de datos utilizada para la gestión y concesión de bloques.

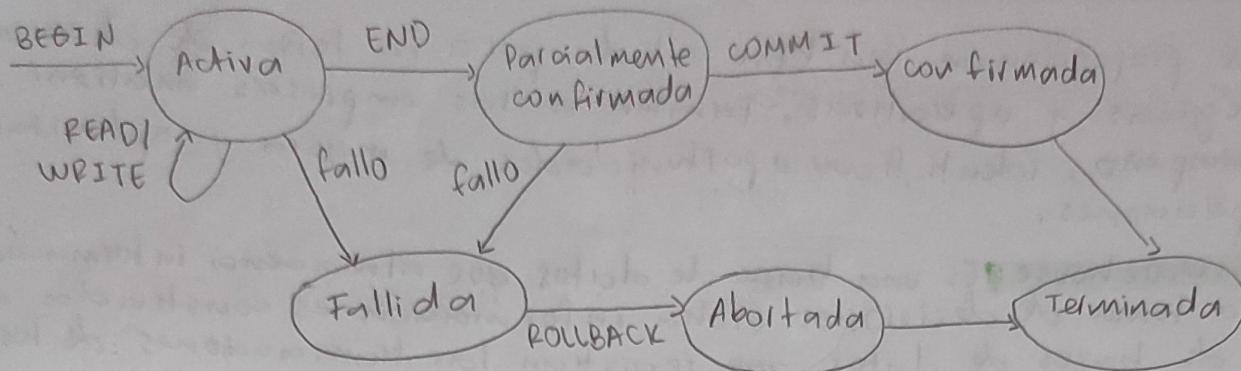
- El gestor de bloques utiliza la siguiente estructura de datos: para cada elemento de datos que está actualmente bloqueado, mantiene una lista enlazada de registros, uno para cada solicitud, en el orden en el que llegaron las solicitudes.

Utiliza una tabla de asociación, indexada por el nombre del elemento de datos, para encontrar la lista enlazada (si la hay) para cada elemento de datos; esta tabla se denomina tabla de bloques. Cada registro de la lista enlazada de cada elemento de datos anota qué transacción hizo la solicitud, y qué modo de bloqueo se solicitó. El registro también anota si la solicitud ya se ha concedido.

19) Defina el concepto de Transacción en SEBDs.

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

Q) Indique y defina las fases del ciclo de vida de una transacción.



Las operaciones hacen que una transacción cambie de un estado a otro. Los estados son los siguientes:

- **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.
- **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.
- **Confirmada:** punto de confirmación y conclusión de la ejecución.
- **Fallida:** si uno de los chequeos falla o si es abortada durante su estado activo.
- **Abortada:** se deshacen los efectos de sus operaciones de escritura sobre la BD.
- **Terminada:** pasa a este estado cuando la transacción desaparece del sistema.

21) a) Por qué se dice que los datos son los nuevos activos de las empresas?

b) Diferencias entre Datawarehouse y Data Mart.

a. Porque proporcionan una base sólida para la toma de decisiones estratégicas y operativas. Permiten a las empresas analizar su desempeño, identificar oportunidades de mejora y optimizar sus operaciones.

b. • **Datawarehouse:** Es una base de datos que almacena información para la toma de decisiones. Dicha información es construida a partir de bases de datos que registran las transacciones de los negocios de la organización.

• **Datamart:** subconjunto de un almacén de datos, generalmente en forma de estrella o copo de nieve.

22) En el modelado dimensional, ¿cuál es la diferencia entre el esquema estrella y los cubos OLAP?

Cite al menos dos consideraciones de implementación de los cubos OLAP.