

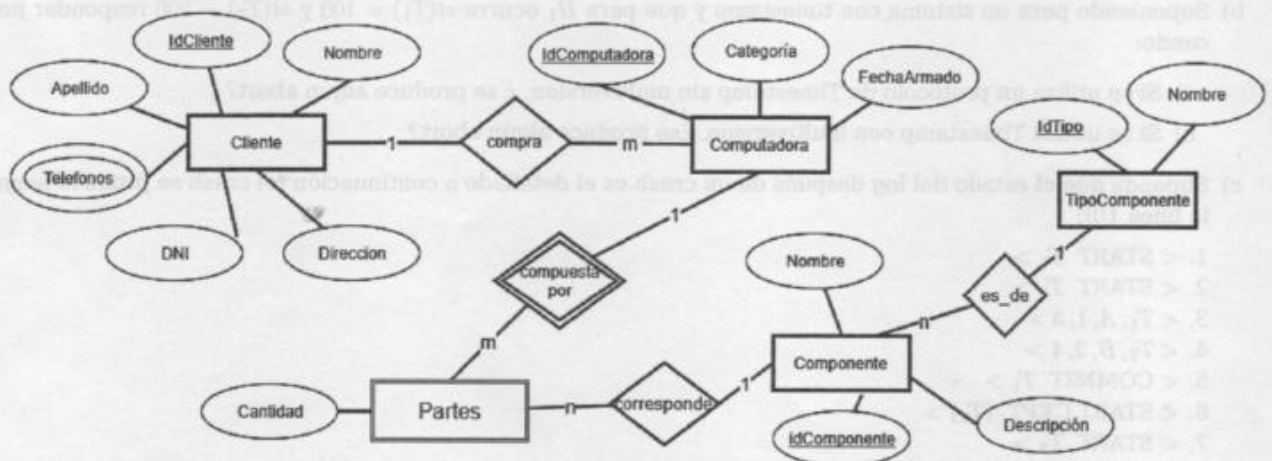
2do Recuperatorio - 1er Cuat. 2017 - Bases de Datos 5/7/2017

- Cada ejercicio debe realizarse en hojas separadas y numeradas. Debe identificarse cada hoja con **nombre, apellido y LU**. Poner el **Nº de Orden** en la primer hoja de CADA EJERCICIO
- La interpretación del enunciado forma parte de la Evaluación.
- Cada tema tiene un criterio de aprobación propio y se recupera por tema. **NO DEBEN COMETERSE ERRORES CONCEPTUALES GRAVES**.
- La devolución de los exámenes corregidos es personal. Los pedidos de revisión se realizarán por escrito, antes de retirar el examen corregido del aula.

1 NOSQL

Criterio de aprobación: se aprueba con 6.5 pts. **a) 4 pts, b) 2 pts c) 2pts d) 2pts.** Errores conceptuales graves o falta de justificación implican cero puntos.

Dado el DER de la figura, donde se modela los datos que necesita una empresa que vende computadoras a medida.



a) Realizar el diseño para una base de datos **document based**: dibujar el **diagrama de interrelación de documentos, justificando las elecciones realizadas**. Especificar al menos un tipo de documento que tenga algún anidamiento con **JSONSchema**.

Tomar en cuenta que las consultas más comunes implican saber los componentes de una computadora, dado un cliente las fechas en las que compro computadoras y las categorías de las mismas.

b) Realizar los diagramas de **Chebotko, justificando**, para una base **column family** que sirva para responder a las siguientes consultas:

i) Dado una computadora y un rango de fechas cuales son los nombres y apellidos de los clientes que compraron alguna computadora en ese rango de fechas.

ii) Todos los clientes ordenados por Nombre y Apellido

c) Diseñar una base **key-value** que soporte el acceso online de los clientes. Cada cliente ingresaría con su DNI y una clave. Y una vez ingresado podría consultar todas las computadoras que compró mediante una lista de **categorías-fecha de armado**. Luego clickeando en cualquier categoría accedería el detalle de la computadora (es decir como esta compuesta la misma). **Justificar. Entregar Diagrama.**

d) Realice una consulta map-reduce que devuelva para cada componente la cantidad de computadoras que están en garantía y que lo contienen. Una computadora está en garantía si la fecha de armado no es mayor a seis meses de antigüedad. Cada documento que recibe como entrada del map contiene los datos de la computadora y una lista de partes que tienen la cantidad y nombre del componente.

2 Concurrency and Recoverability

Criterio de aprobación: se aprueba con 6.5 pts. Valores parciales: a) 4, b) 3, c) 3. Debe tener **al menos algo correcto en cada ítem**

a) Dadas las siguientes historias sobre las transacciones T_1 y T_2 y la siguiente historia:

$H_1 : r_2(X), r_1(X), w_1(X), c_1, w_2(X), r_2(Y), w_2(Y), c_2$

$H_2 : r_2(X), w_2(X), r_2(Y), r_1(X), w_2(Y), w_1(X), c_1, c_2$

Se pide

- Construir el $SG(H_1)$ y $SG(H_2)$ e indicar si son SR. En caso afirmativo indicar las historias seriales equivalentes.
- Clasificar H_1 y H_2 con respecto a los criterios de **Recuperabilidad**.
- ¿ H_1 y H_2 son historias equivalentes? ¿Por qué?
- Reescribir H_2 en el modelo de **Locking Ternario** sin *upgrade lock*. ¿La historia resultante respeta el protocolo 2PL?

Justifique todas sus respuestas

b) Suponiendo para un sistema con timestamp y que para H_1 ocurre $st(T_1) = 100$ y $st(T_2) = 200$ responder justificando:

- Si se utiliza un protocolo de Timestamp sin multiversion, ¿se produce algún abort?
- Si se utiliza Timestamp con multiversion, ¿se produce algún abort?

c) Suponga que el estado del log después de un crash es el detallado a continuación (el crash se produce luego de la línea 10):

- < START T_1 >
- < START T_2 >
- < $T_1, A, 1, 3$ >
- < $T_2, B, 2, 4$ >
- < COMMIT T_1 >
- < START CKPT (T_2) >
- < START T_3 >
- < $T_3, C, 8, 10$ >
- < COMMIT T_3 >
- < END CKPT >

Asumiendo un mecanismo de recuperación Undo/Redo con checkpoints no quiescente especifique claramente, justificando las respuestas:

- Las transacciones a deshacer y rehacer.
- ¿Qué valor tiene grabado en el disco el elemento B antes del recovery?
- Los cambios efectuados en la base de datos.
- Los cambios efectuados en el log.

3 Optimización

Criterio de aprobación: se aprueba con 6.5 pts, sin errores conceptuales, y **el plan de ejecución propuesto debe ser muy cercano al plan de menor costo posible.**

Se tiene una Base de Datos con el siguiente esquema:

Estudio (idEstudio, nombreEstudio, pais)

Pelicula (idPelicula, nombre, año, genero, idEstudio, idDirector)

Director (idDirector, nombreApellido, nacionalidad)

Los Estudios producen muchas películas. Las películas son dirigidas por un director y producidas por un estudio. Los directores pueden dirigir muchas películas

Se desea optimizar la siguiente consulta:

```
SELECT D.nombreApellido, P.Nombre, P.Año,
FROM Estudio E , Pelicula P, Director D
WHERE E.idEstudio = P.idEstudio AND
      P.idDirector = D.idDirector AND
      P.año >= 2010 AND P.año <= 2012 AND
      P.Genero = "Policial" AND
      E.Pais = "Argentina"
```

Se tienen la siguiente información sobre índices y tamaños:

- Índices (altura de árboles B+ es $X = 3$):
 - I_1 índice **hash** sobre $D.idDirector$
 - I_2 índice B+ **clustered** sobre $E.pais$
 - I_3 índice B+ **clustered** sobre $P.Genero$
 - I_4 índice B+ **unclustered** sobre $P.Año$
 - No hay dos directores con el mismo nombre.
 - Hay 10.000 películas policiales.
 - Se producen 200 películas por año.
 - $I_{Estudio,pais} = 20$, $T_{Pelicula} = 350.000$, $T_{Director} = 100.000$, $T_{Estudios} = 3.000$
 - Tamaño de los bloques: $LB = 2048$ bytes. Tamaño de todos los atributos 32 bytes.
 - La cantidad de bloques disponibles de memoria es $B = 102$.
 - Suponga distribución uniforme. Asumir que los punteros a tupla que se necesiten recorrer en los índices B+ entran en una hoja.
- a) Escriba el árbol canónico y partir de él modifíquelo para obtener un plan de ejecución optimizado indicando **solamente el árbol final**. Justifique cada una de las decisiones que tome para su propuesta de plan de ejecución. Justifique también porque decide utilizar o no utilizar cada uno de los índices disponibles.

b) Calcule el costo del árbol optimizado.