

IIC 2413 – Bases de Datos
 Interrogación 3

Pregunta 1: External Merge Sort y Transacciones

a) [2 pts] Considere una relación de 3.000.000 páginas. Indique el número de fases y de I/O del External Merge Sort optimizado visto en clases cuando:

- En Buffer caben 11 páginas.
- En Buffer caben 101 páginas.

Comente sus resultados (puede dejar expresados los logaritmos). Además diga a cuanto converge el número de I/O cuando el buffer sigue creciendo. Indique que pasa si se evita escribir el resultado de la última fase y el resultado se imprime directamente en consola.

b) [4 pts] Considere el Schedule del cuadro 1. Diga si es o no *conflict serializable*. En caso de que no lo sea, explique por qué e indique cómo Strict-2PL puede volverse un problema si las transacciones llegan en ese orden.

| T1 | T2 | T3 |
|------|------|--------------|
| R(a) | R(b) | W(a) W(c) |
| R(c) | | |
| | R(c) | |

Cuadro 1: schedule problema 2.

Pregunta 2: Logging

Undo Logging

Suponga que su sistema tuvo una falla. Al reiniciar el sistema, el sistema se encuentra con el *log file* que se muestra a continuación, en la tabla “Log Undo”. Suponiendo que la política de *recovery* es la de *Undo Logging*, indique:

- Hasta qué parte del *log* debo leer.
- Qué variables deben deshacer sus cambios y cuál es el valor con el que quedarán.
- Qué variables (de las que aparecen en el *log*) no son cambiadas en el proceso.

| Log Undo |
|-------------------|
| <START T1> |
| <START T2> |
| <T1, a, 4> |
| <T2, b, 5> |
| <T2, c, 10> |
| <COMMIT T1> |
| <START CKPT (T2)> |
| <START T3> |
| <START T4> |
| <T3, a, 10> |
| <T2, b, 7> |
| <T4, d, 5> |
| <COMMIT T2> |
| <END CKPT> |
| <START T5> |
| <COMMIT T3> |
| <T5, e, -3> |

Redo Logging

Suponga que su sistema tuvo una falla. Al reiniciar el sistema, el sistema se encuentra con el *log file* que se muestra a continuación, en la tabla “Log Redo”. Suponiendo que la política de *recovery* es la de *Redo Logging*, indique:

- Desde qué parte del *log* debo comenzar el proceso de *redo*.
- Qué variables deben rehacer sus cambios y cuál es el valor con el que quedarán.
- Qué variables (de las que aparecen en el *log*) no son cambiadas en el proceso.
- Si no hubiesemos encontrado la línea <END CKPT>, ¿desde qué parte del *log* debería comenzar el proceso de *redo*?

| Log Redo |
|-------------------|
| <START T1> |
| <T1, a, 1> |
| <COMMIT T1> |
| <START T2> |
| <T2, b, 2> |
| <T2, c, 3> |
| <COMMIT T2> |
| <START T3> |
| <T3, a, 10> |
| <START CKPT (T3)> |
| <T3, d, 23> |
| <START T4> |
| <END CKPT> |
| <COMMIT T3> |
| <T4, e, 11> |

Pregunta 3: MongoDB

¡Felicitaciones! Fuiste recientemente contratado en la red social *Fotograma*¹. Esta red social tiene utiliza una base de datos MongoDB que posee un modelo bastante simple:

- Existe una colección de usuarios con datos básicos que toda red social debe tener.
- Existe una colección de fotogramas. Estas son las imágenes que los usuarios suben a la red social. Entre otros atributos, se destaca la presencia de la URL de la imagen y la descripción, que corresponde a un texto.
- Cada documento que corresponde a la colección de usuarios tiene una lista con los id de los fotogramas a los que ha marcado con “me gusta”.

Un ejemplo de documentos que se pueden encontrar en la base de datos MongoDB son los siguientes:

```
// Usuarios
{ "uid": 23,
  "nombre": "Fernando Pieressa",
  "bio": {
    "estudia_en": "PUC",
    "edad": 23,
    "descr": "Link a mi último vídeo: yu.tuv/video1234"
  },
  "me_gusta": [3, 5, 6] }

// Fotogramas
{ "fid": 1,
  "posteada_por": 23,
  "fotograma": "fotograma.com/foto/12345.jpg",
  "descripción": "Foto de mi preparación para subir el Cerro El Plomo #Trekking #Outdoor"}
```

En tu primer día de empresa te encargaron diseñar algunas consultas. Dado que en el curso de bases de datos aprendiste MongoDB a la perfección, esto no va a ser un problema para ti.

Utilizando PyMongo o JavaScript se pide un procedimiento para cada una de las consultas a continuación. Si necesitas un índice por texto debes indicar el comando para crearlo.

- **[1.5 pts]** Entregue el número de “Me gusta” que suman todos los fotogramas que en su descripción tengan el texto “#Trekking” y que no tengan el texto “Cerro La Cruz”.
- **[2 pts]** Imprima el identificador de cada usuario seguido de las URL de todos sus fotogramas.
- **[2.5 pts]** Imprima el identificador de cada usuario junto al número total de “Me gusta” que ha recibido. Esto es, sumar los “Me gusta” de cada uno de sus fotogramas.

Bonus [0.6 pts] Mencione 6 cerros que pertenecen a la Sierra de San Ramón.

¹Sí, es una mezcla de Fotolog con Instagram.