
Control 3:**Gerenciador de Almacenamiento**Docente: [Ana Maria Cuadros](#) Valdivia

05 de Mayo del 2024

Apellidos y Nombres: _____

Grupo: _____

1. Para realizar un seguimiento de la cantidad de espacio libre en una página, se utiliza el directorio de páginas?
2. Para el siguiente esquema, cuál es el tamaño más pequeño, en bytes, que se puede tener para un registro, si el encabezado del registro es de 5 bytes?
(booleano = 1 byte, fecha = 8 bytes)

name VARCHAR
student BOOLEAN
birthday DATE
state VARCHAR

3. Para celebrar el final del semestre, los alumnos del curso de BD deciden participar en un sorteo. Cada vez que un alumno compra un boleto, se inserta una tupla que contiene la primera letra de su nombre y el número de su boleto en un *file* no ordenado. Tenga en cuenta que cada número del ticket de lotería es único. Al final del día el *file* contiene las siguientes tuplas:

Name (Char(1))	Number (Integer)
E	2245
Y	2240
V	2235
E	2243
Y	2244
E	0007
R	2241
C	2242
S	2250
L	2249

Después de un tiempo, el proceso de sorteo dio origen a la compañía Lottery. En un día cualquiera, esperan que más de 10,000 personas compren tickets de lotería. Al comienzo de cda día, Loterry crea un nuevo archivo. Cada vez

que una persona compra un boleto, se inserta un nuevo registro (con el esquema mostrado anteriormente) en el *file*. Al final de cada día, se elige un número ganador y el registro correspondiente se saca del *file*.

Dada esta especificación, marque las letras de los parámetros (una en cada columna) que debería optimizar el desempeño para que Loterry realice sus tareas:

Page Format	File Layout	Index
(a) Fixed length (Packed)	(d) Heap File	(f) Clustered Index on Number
(b) Fixed length (Unpacked)	(e) Sorted File	(g) Unclustered Index on Number
(c) Slotted Page		(h) None

4.

You are working as a database administrator (DBA) at ClikClak, an anonymous text message forum app. Their data is stored in a single table with the following schema:

`Post (postid int, userid int, pdate int, content text)`

where `postid` is a primary key for `Post`. `Post` takes up $N = 10,000$ pages on disk.

The `Post` table is partitioned across three servers based on the data's access pattern.

Server ID	Read/Write Ratio	% of N	Access Pattern Details
S1	80/20	20%	sequential scans in chronological order
S2	40/60	60%	posts made by a user, shown on their profile
S3	99/1	20%	queries for random specific posts no temporal locality

2. **(12 points)** For each server, fill in the blank on your answer sheet to choose parameters that best optimize performance for that server's access pattern. Don't ignore the maintenance cost of insertions/updates when considering performance.

- 2b. Select a file layout and index:
- A. HeapFile, Unclustered B+ tree on postid
 - B. HeapFile, Clustered B+ tree on userid
 - C. SortedFile on date, Unclustered B+ tree on date
 - D. SortedFile on date, No index
 - E. SortedFile on userid, Unclustered B+ tree on content
 - F. SortedFile on userid, No index

Rpta:

1. Verdadero
2. The smallest size of the VLR is 14 bytes and occurs when both name and state are null.
 $5 \text{ (record header)} + 1 \text{ (boolean)} + 8 \text{ (date)} = 14$
3. Solution: a, d, h.

The Lottery Company is doing many inserts, one equality search and no deletes. So we should primarily focus on optimizing inserts.

A fixed length (packed) page is the best page layout because records have a fixed length of 5 bytes (1 char, 1 int), and since we don't have to worry about deletes.

A heap file is the best file layout because many inserts are being done and heap files have a constant time insert while the cost to maintain the sorted file becomes progressively worse.

Building any type of index will add unnecessary overhead to our inserts so it is best to build none.

4. S1: D - SortedFile on date, no index
 - just do sequential scans on sorted file
- S2: B - HeapFile with Clustered B+ tree on userid
 - since this constitutes 60% of the data, and fairly write heavy, clustered B+ tree is better choice than SortedFile
- S3: A - heapfile, Unclustered B+ tree on postid
 - use unclustered for point lookups, and no need for sorted file

