

CSci 4707
Homework 3
Spring 2015

Chapter 8 and 12
Due Thursday, 04/09/2015 14:30

A. (5 Points) Submission Guidelines

- All homeworks **must be typed**
- All homeworks due at the **beginning of the class in paper**
- **No partial credit awarded for this section**

B. (50 Points) Chapter 8

1. **(20 Points, 5 Points each)** Assume you have a buffer with a maximum capacity of 5 pages. Below is a queue of the current state of the buffer with its current pin count with the time when pin counts becomes 0 (the pages were inserted sequentially: Page A comes before Page B, etc. and time 0 came first before time 1 etc.).

A:1:NaN	B:0:2	C:0:3	D:1:NaN	E:0:1
---------	-------	-------	---------	-------

Now, there are 4 pages (F, G, H, and I) need to be inserted into the buffer. For each buffer management policy below, show the result of the buffer after the insertion. Just show the Page name on each buffer slot.

- a. **(5 Points)** FIFO (First In First Out)
 - b. **(5 Points)** MRU (Most Recently Used)
 - c. **(5 Points)** LRU (Least Recently Used)
 - d. **(5 Points)** Clock (The clock is currently pointing to C)
2. **(5 Points)** The file systems of modern operating systems also buffer data from disk in-memory for fast access, and they have very sophisticated

replacement strategies. Why does a database system manage its own buffer pool rather than letting the operating system manage the buffer? Give two reasons in one sentence each.

3. **(10 Points, 5 Points each)** Consider a file that is stored on disk. What buffer manager replacement policies would you choose for each of the following file access patterns? Clearly state the reason for your choice; answers without a correct reason will receive zero points.
- a. **(5 Points)** The entire file is sequentially scanned many times.
 - b. **(5 Points)** File pages are accessed according to a Zipfian distribution, i.e., popular items are requested very frequently, and unpopular items are requested rarely.
4. **(15 Points, 5 Points each)** Answer the following questions:
- a. **(5 Points)** When will equality search is better using Unclustered B+Tree rather than Sorted?
 - b. **(5 Points)** Is it possible to have an Unclustered B+Tree index with Alternative 1? If yes, what is the advantage of using Alternative 1 compared to Alternative 2 or 3? If no, why not?
 - c. **(5 Points)** What is the main difference between B-Tree (Not Binary Tree) and B+-Tree?

C. **(45 Points) Chapter 12**

1. **(20 Points, 5 Points each)** Consider a relation with this schema: Companies(cid:integer, cname:string, age:integer, state:string, category:string)
- Suppose that the following indexes, all using Alternative (2) from your textbook for data entries, exist: a hash index on cid, an unclustered B+Tree index on age, a hash index on state and a clustered B+Tree index on <state, age>. What is the most selective access path for retrieving all Company tuples that satisfy that condition (only):
- a. **(5 Points)** age > 10

- b. (5 Points)** state = "CA"
 - c. (5 Points)** age > 5 AND category = "Technology"
 - d. (5 Points)** state = "NY" AND age > 5 AND category = "Finance"
- 2. (5 Points)** Suppose that, for the above selection condition (b), you want to compute the average age for each State (i.e., group by state). For selection condition (b), describe the least expensive evaluation method.
- 3. (20 Points, 4 Points each)** Suppose relation R has 1000 records and relation S has 5000 records. One page can contain 10 R records or 500 S records. What are the cost of the following join operation between R and S?
 - a. (4 Points)** Index Nested Loop Join with Clustered B+Tree Index on S (R is the outer).
 - b. (4 Points)** Index Nested Loop Join with Clustered Hash Index on R (S is the outer).
 - c. (4 Points)** Index Nested Loop Join with Unclustered B+Tree Index on S (R is the outer).
 - d. (4 Points)** Tuple Nested Loop Join (S is the outer).