CSE 232A Fall 2019 Midterm Exam 1
Answers

Part 1. [5pts] For each statement below, indicate if it is (A) True or (B) False.

1. A key reason for RDBMSs to manage their own buffer pools versus letting the OS do it is to reduce query latency by exploiting data access patterns of relational queries.
   True

2. In the unpacked layout for fixed-length records, deleting a record will alter the record IDs of other records on the same page.
   False

3. Column stores were motivated by the need to reduce query latency for primarily read-only aggregate queries.
   True

4. An RDBMS typically stores its catalog as a relational database itself.
   True

5. The main benefit of learned indexes over B+ tree indexes are that the former usually have orders of magnitude lower query latency.
   False

Part 2. [6pts] For each question below, select the right option (only one is correct).

6. Which of these components in a storage hierarchy is not persistent?
   (A) Tape   (B) Hard disks   (C) Flash / SSDs   (D) NVM   (E) DRAM
   ANSWER: (E)

7. Which of these is not a basic operator in relational algebra?
   (A) \(\bowtie\)   (B) \(\sigma\)   (C) \(\pi\)   (D) \(\times\)   (E) \(\rho\)
   ANSWER: (A)
8. You are given the relation schema Movies (MID, MName, Year, Director) with primary key MID. Which of the following indexes does not match the following query: \( \sigma_{Year=2018}(Movies) \)?

(A) Clustered B+ tree index on Year
(B) Clustered B+ tree index on Year, Director
(C) Hash index on Year
(D) Hash index on Year, Director
(E) Clustered B+ tree index on Year, MID

**ANSWER: (D)**

---

**Part 3. [9pts]** Suppose we are sorting a relation with **20 million** pages, and we have **1000** buffer pages for the external merge sort (EMS). A “pass” is defined as one read and write of the whole file. For each of the following three questions, include both the sort and merge phases.

9. **[3pts]** How many passes will a multi-way EMS perform, assuming we use replacement sort for internal sorting but no blocked I/O or double buffering?

   (A) 2   (B) 3   (C) 4   (D) 5   (E) 6

   **ANSWER: (B) 3.** \( N' = 10,000; F = 999. \)

10. **[3pts]** How many passes will a multi-way EMS perform, assuming we do not use replacement sort and do not use blocked I/O but use double buffering?

   (A) 2   (B) 3   (C) 4   (D) 5   (E) 6

   **ANSWER: (B) 3.** \( N' = 20,000; F = 499. \)

11. **[3pts]** How many passes will a multi-way EMS perform, assuming we do not use replacement sort and do not use double buffering but use blocked I/O with block sizes of **100** pages?

   (A) 2   (B) 3   (C) 4   (D) 5   (E) 6

   **ANSWER: (E) 6.** \( N' = 20,000; F = 9. \)
Part 4. [10pts] Consider the following extendible hash index with 4 slots per bucket.

12. [3pts] What is the global depth after the following sequence of operations finish successfully: insert 44*, delete 20*, and insert 67*?
   (A) 1    (B) 2    (C) 3    (D) 4    (E) 0

   ANSWER: (B) 2. Bucket B splits but its LD is 1; so GD is unaffected.

13. [3pts] Consider the same given index (i.e., ignore the above question). What is the maximum total number of entries that can be successfully inserted before any page must split?
   (A) 1    (B) 2    (C) 3    (D) 4    (E) 5

   ANSWER: (C) 3. Total number of free slots in the buckets.

14. [4pts] Consider the same given index (i.e., ignore the above questions). What is the minimum total number of entries that must be deleted after which global depth becomes 1?
   (A) 1    (B) 4    (C) 5    (D) 8    (E) 9

   ANSWER: (A) 1. Just delete 16*.

Part 5. [10pts] Consider the following B+ tree index that uses the AltRID alternative. You are given that its order is 2. Each leaf or non-leaf page can hold at most 4 entries.
15. **[3pts]** What is the *maximum* number of entries that can be successfully inserted *before* any page must split?

(A) 6  (B) 8  (C) 10  (D) 11  (E) 12

**ANSWER:** (D) 11. Total number of free slots in the leaf pages.

16. **[3pts]** Consider the same given index (i.e., ignore the above question). What is the *minimum* total number of entries that must be successfully deleted *before* any page must merge?

(A) 1  (B) 2  (C) 3  (D) 4  (E) 5

**ANSWER:** (A) 1. Just make any leaf pages with 2 entries underflow.

17. **[4pts]** Consider the same given index (i.e., ignore the above questions). Suppose a delete operation is executed with the predicate $\text{SearchKey} = 31$. What is the total number of pages in the index after this operation finishes successfully?

(A) 5  (B) 6  (C) 7  (D) 8  (E) 9

**ANSWER:** (B) 6. Two leaf pages merge, triggering a recursive merge and root pull-down.

**Part 6. [10pts]** Who does not go shopping? Given the following simplified relational database schema for shopping, translate each natural language query or SQL query into an equivalent relational algebra query (only one option is correct).

- **Products** (PID, PName, Vendor, Price)
- **Customers** (CID, CName, Zipcode, Age)
- **Purchases** (CID, PID, Timestamp, Quantity)
Products.PID and Customers.CID are primary keys in their respective tables. In Purchases, CID and PID are foreign keys referring to Customers.CID and Products.PID, respectively.

18. [2pts] Which vendors sell products priced over 50?

(A) $\sigma_{\text{Vendor}, \text{Price}>50}(\pi(\text{Products}))$

(B) $\pi_{\text{Vendor}, \text{Price}}(\sigma_{\text{Price}>50}(\text{Products}))$

(C) $\sigma_{\text{Price}>50}(\pi_{\text{Vendor}}(\text{Products}))$

(D) $\pi_{\text{Vendor}}(\gamma_{\text{Price}>50}(\text{Products}))$

(E) $\pi_{\text{Vendor}}(\sigma_{\text{Price}>50}(\text{Products}))$

ANSWER: (E)

19. [3pts] What is the average age of customers who bought the product with PID 123?

(A) $\pi_{\text{Age}}(\sigma_{\text{PID}=123}(\text{Purchases}) \bowtie \text{Customers})$

(B) $\pi_{\text{CID}, \text{Age}}(\text{Customers} \bowtie \sigma_{\text{PID}=123}(\text{Purchases}))$

(C) $\gamma_{\text{AVG(Age)}}(\text{Customers} \bowtie \sigma_{\text{PID}=123}(\text{Purchases}))$

(D) $\gamma_{\text{AVG(Age)}}(\pi_{\text{CID}, \text{Age}}(\sigma_{\text{PID}=123}(\text{Purchases}) \bowtie \text{Customers}))$

(E) $\gamma_{\text{CID}, \text{AVG(Age)}}(\pi_{\text{CID}}(\sigma_{\text{PID}=123}(\text{Purchases}) \bowtie \text{Customers}))$

ANSWER: (D)
20. [5pts] SELECT DISTINCT C.CID, C.CName
    FROM Customers C, Purchases P
    WHERE C.CID = P.CID AND P.PID IN
    (SELECT DISTINCT PID
    FROM Purchases P2, Customers C2
    WHERE C2.CName = 'Alice')

For brevity, we use C for Customers and P for Purchases (rename).

(A) $\pi_{CID,CName}(\sigma_{CName='Alice'}(C) \bowtie P \bowtie (\pi_{PID}(C \bowtie P)))$

(B) $\pi_{CID,CName}(C \bowtie P \bowtie (\pi_{PID}(\sigma_{CName='Alice'}(C) \bowtie P)))$

(C) $\pi_{CID,CName}(\sigma_{CName='Alice'}(C) \bowtie \pi_{PID}(C \bowtie P) \bowtie P)$

(D) $\pi_{CID,CName}(\sigma_{CName='Alice'}(C) \bowtie \pi_{C2.CID,P.PID}(\rho_{C2}(C) \bowtie P) \bowtie P)$

(E) $\pi_{CID,CName}(C \bowtie P \bowtie (\pi_{CID}(\sigma_{CName='Alice'}(C) \bowtie P)))$

**Answer:** Unfortunately, I missed the join condition in the nested subquery. So, everyone will be given these 5pts. Had the join condition been specified, the correct answer would have been (B).