Name:

CS 33007 Introduction to Database System Design, Fall 2018

Final Exam Answer Key

Instructions:

- *This examination is closed book (no access to book, lecture notes, phone, laptop, tablet etc.). But you can have notes on* ***one side of a regular paper****.*
- *Please write your answer in the given blank space for each question. If your answer doesn't fit in the given space, you can use back side of the papers but write question number.*

Total Points:100 **Time:** 12:45 PM – 3PM

1. What are the goals of database normalizations? What are the differences between BCNF and 3NF.

[15 points]

Answer:

In database normalization, we want to decompose bad relation schemas such a way that,

- each relation schema is in good form after decomposition
- the decomposition is a lossless-join
- **Preferably, the decomposition is dependency preserving.**

Difference between BCNF and 3NF,

1. BCNF doesn't allow redundancy in database where 3NF does.

2. If any relation schema is decomposed in 3NF, it guarantees functional dependency preservation. This is not true for BCNF.

2. Identify possible functional dependencies from the following schema of car-sell relation used by a Toyota dealer, **[10 points]**

Car-sell (transactionID, customerID, CustomerName, customerAddress, carModel, MakeYear, carPrice, NoOfSeats)

Answer:

i) transactionID → customerID, CustomerName, customerAddress, carModel, MakeYear, carPrice, NoOfSeats

ii) customerID → CustomerName, customerAddress

iii) carModel, MakeYear → carPrice, NoOfSeats

3. Suppose we have a set of functional dependency F = {A → BC, CD → E, B → D, E → A} on the schema R = (A, B, C, D, E). Show that the decomposition of R into R1= (A, B, C) and R2 = (A, D, E) is lossless-join decomposition. **[10 points]**

Answer:

From Attribute closure set algorithms (A)+ = ABCDE

So, A is a candidate key, i.e. A - >ABCDE. Since R1 ∩ R2 = A, therefore R1 ∩ R2 → R1. So, we can say the decomposition of R into R1 and R2 is lossless – join decomposition.

4. Discuss file organization for variable size records with an example. **[15 points]**

Answer:

Please see slide no 46 in chapter 10

5. *Map the instructor table given below in four buckets using the hash function H(K), where K is the instructor ID and hash function H is defined as summation of digits in ID modulo 4. Considering each bucket can store only 4 records, use overflow bucket if necessary.*

[15 points]

ID	name	dept_name	salary
45565	Katz	Comp. Sci	75000
98345	Kim	Elc. Eng.	80000
58583	Califieri	History	62000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76766	Crick	Biology	72000

Answer:

See Homework6 key, question number 8.

6. Suppose a relation has 1000 records which is stored in a file using B^+ tree index file. Assume every record has unique search key. Each node in the B^+tree contains maximum 6 search keys. What will be the maximum height of the B^+tree**?** **[10 points]**

Max height for B^+tree is $\lceil \log_{\lceil n/2 \rceil}(K) \rceil$, here n = 6+1 = 7 and K = 1000. So Max height is 5.

7. Suppose available buffer size to perform a join operation is 1GB (1024MB). Block size is 10MB. Two tables to be joined are ***students*** and ***instructors***. Table *students* has 500 records which takes 100 blocks to store. Table *instructors* has 100 records and take 400 blocks to store. If you use nested join loop algorithm for joining. **[15 points]**

i. Which table should be used in inner loop and why.

ii. What will be the number of block transfers and seeks.

Answer:

i. In nested loop, all blocks of inner loop table need to be accessed for every record of outer loop table. So inner table is accessed frequently. Since the students table entirely can fit into the buffer, loading whole student table in buffer and using it in inner loop will reduce overall number of block transfers and seeks.

ii. Each table need to be read only once from the disk. So total block transfer = 100+400 =500. Each table need to seek only once and then access whole table sequentially. So, total number of seeks is 2.

8. What is the purpose of query optimization? Summarize the steps of query optimization. **[10 points]**

Answer:

Query Optimization: Amongst all equivalent query evaluation plans choose the one with lowest cost. Cost is estimated using statistical information from the database catalog e.g. number of tuples in each relation, size of tuples, etc.

Steps in cost-based query optimization

1. Generate logically equivalent relational algebra expressions using equivalence rules.
2. Annotate resultant expressions to get alternative query evaluation plans.
3. Choose the cheapest plan based on estimated cost.