CS 33007 Introduction to Database System Design, Summer 2018

Final Exam

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 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:** 3:00PM – 4:30PM

1.

a. Prepare a partial ER diagram for a relational database that will allow Facebook to store information about the users and their posted text. The diagram must list all the necessary attributes of the entity sets and specify their primary keys based on real word practice. **[15 points]**

b. Covert the ER diagram in (a) to a non-redundant set of relation schemas.   **[8 points]**

c. Identify functional dependencies on the obtained schemas.   **[7 points]**

2.

    a. What are the goals of database normalization? When would you choose 3NF over BCNF. **[10 points]**

    b. Given $F = \{A \rightarrow B, A \rightarrow C, CG \rightarrow H, CG \rightarrow I, A \rightarrow G\}$, set of functional dependencies on R (A, B, C, G, H, I), Show that A is a super key of R. **[10 points]**

3.

a. What are the benefits of using index file for a relation? Explain the tradeoff between space overhead and query execution time of using index file. **[10 points]**

b. What are the disadvantages of static hashing? **[8 points]**

c. When is it better to use bitmap indexing? **[7 points]**

4.

a. 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. [**20 points**]

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

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

b. What are the advantages of using pipelining to evaluate relational expressions? [**5 points**]