Homework 3: Relational Database Design Theory (100 points)

Due Date: Wed, Apr 25 (5:00 PM)

Submission

All HW assignments should be turned in with a filename that contains both your student ID and your name (e.g., 12345678_John_Doe.pdf) and must be submitted online, as a PDF file, through the associated (HW3 in this case) HW3 in Gradescope. See the table below for HW 3 submission opportunities. Note that after the last deadline, Thursday Apr 26th, no further HW 3 submissions will be accepted at all. That is, we will not accept assignments after that time since we will be publishing the solution at that time -- this is important so that students who want to can see the solution before the first Midterm exam. Please turn in all of your work on time! If possible, save your one dropped assignment for the end of the term when you are most likely to want/need it.

<table>
<thead>
<tr>
<th>Date / Time</th>
<th>Grade Implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wednesday, Apr 25 (5:00 PM)</td>
<td>Full credit will be available</td>
</tr>
<tr>
<td>Thursday, Apr 26 (5:00PM)</td>
<td>20 points will be deducted</td>
</tr>
</tbody>
</table>

Submission note: please use the provided template to answer each question. Normally, your solution for each question should not EXCEED 1 page for grading purposes. Since there are four questions, four pages will be enough. However, if you think you can’t answer one question in a single page, you can insert one additional blank page and use it to answer a question in two pages. In any case, please always start a new question at the beginning of a page. (The template ensures this if you don’t exceed one page per question.)
1. [25 pts]
On further examination of your existing relational schema, your boss is troubled by the fact that two users can have the same phone number as one of their numbers but with different types (e.g., one could say it’s a mobile phone and the other could say it’s a home phone). She wants to make sure that when two users do have a phone number in common, the type information is consistent.

(a) [5 pts] Start by looking at the existing Phone table. What non-trivial functional dependencies does the current table have?

(b) [5 pts] What additional functional dependency is needed to express your boss' rule about phone type consistency?

(c) [5 pts] Given this new functional dependency, what is the new primary key for the Phone table given its current schema?

(d) [5 pts] What normal form is the Phone table currently in after adding the boss’s new rule? Briefly show your reasoning.

(e) [5 pts] Decompose Phone into multiple tables to produce a well-normalized design (BCNF if possible, 3NF if not) if the current design isn't already there. Your solution must be lossless join and dependency preserving. List the decomposed tables along with their candidate keys.
2. [25 pts]
After the initial development of Peeeza, your boss wants to add a new feature to keep track of the post viewing activities of students. This will allow the teaching staff to see how much time students generally spend on the system and which posts are more popular among students. An initial schema design as well as its description and some sample data are shown below. (For simplicity, foreign key references were omitted.) Please answer the following questions.

-- New table for recording post viewing activities
-- Each time a user (user_id) views a post (post_id), a record is inserted into this table
-- The attribute view_time is when the viewing happens
-- The attribute staying_time is the total time (in seconds) for which the user viewed the post
-- A user can potentially view multiple posts at the same time (view_time); however, if a user opens multiple browser tabs on the exact same post at the exact same time, Peeeza will detect that case and record it as one viewing of the post at that time.

CREATE TABLE User_Post_Views(
    user_id VARCHAR(20) NOT NULL,
    post_id VARCHAR(20) NOT NULL,
    view_time DATETIME NOT NULL,
    staying_time INTEGER);

INSERT INTO User_Post_Views(user_id, post_id, view_time, staying_time)
VALUES
    ('1', '1', '2018-04-21 05:06:10', 60),
    ('1', '2', '2018-04-21 05:06:10', 20),
    ('1', '1', '2018-04-27 09:25:10', 25);

(a) [5pts] List all of the functional dependencies involving the attributes of the relation in question. (Just give the basic list, not the closure of that list. ☕)

(b) [5 pts] What are the candidate keys for this relation?

c) [5 pts] Does the relation satisfy 1NF [Yes/No]? Given a short reasoning (<= 2 sentences).

d) [5 pts] Does the relation satisfy 2NF [Yes/No]? Given a short reasoning (<= 2 sentences).

e) [5 pts] Does the relation satisfy 3NF [Yes/No]? Given a short reasoning (<= 2 sentences).
3. [25 pts]
Given the relation R and the following functional dependencies, answer the following questions.

R(A, B, C, D, E)
All attributes contain only atomic values.

FD1: A → BC
FD2: CD → E
FD3: B → D
FD4: A → E

(a) [5 pts] Compute A+, the attribute closure of attribute A.

(b) [5 pts] List the candidate keys of R.

(c) [5 pts] What’s the highest normal form that R satisfies and why?

(d) [5 pts] If R is not already at least in 3NF, then normalize R into 3NF and show the resulting relation(s) and their candidate keys. Your decomposition should be both join-lossless and dependency-preserving. If R is already in 3NF, just list the candidate keys of R.

(e) [5 pts] Is your answer to (d) also in BCNF? Why or why not?
4. [25 pts]
Given the relation R and the following functional dependencies, answer the following questions.

R(A, B, C, D, E, F, G)
All attributes contain only atomic values.
FD1: AB → C
FD2: BC → E
FD3: B → D
FD4: F → G

(a) [5 pts] Compute A+, the attribute closure of attribute A.

(b) [5 pts] List the candidate keys of R.

(c) [5 pts] What’s the highest normal form that R satisfies and why?

(d) [5 pts] If R is not already at least in 3NF, then normalize R into 3NF and show the resulting relation(s) and their candidate keys. Your decomposition should be both join-lossless and dependency-preserving. If R is already in 3NF, just list the candidate keys of R.

(e) [5 pts] Is your answer to (d) also in BCNF? Why or why not?