Corrections - CS345 Exam 2 Fall 2023 **Name Jax Lubkowitz**

Some correct answers are better than other correct answers. Full credit for the best answer. Show work to receive partial credit.

Corrections: For *some* points back (I am not guaranteeing how many, but no more than half of what was deducted) make corrections directly to this Word Doc. Do not use pen/pencil. Turn in a stapled, double-sided printout. Use proper notation, use proper symbols. For example, for an arrow use → and not ->. Make sure to turn in your old exam separately (not stapled to the corrections) because I need to know how many points I deducted. If you are not making modifications to a question, then leave it blank. Do not redo questions that you got correct. Feel free to add as much space as needed between questions and to delete extra whitespace. **Put your answers in dark blue.**

1. [5] State and briefly explain the two primary goals of database design.

The two goals of database design are to eliminate/reduce redundancy and be complete. Reducing redundancy eliminates update anomalies i.e. changes in one row's values, don't force another change in another row that correspond to the same value to also be changed (otherwise the database becomes inconsistent). Completeness on the other hand allows for rows in a database to be inserted or deleted without losing information or gaining new information. For example, in an insertion anomaly on a school database, we might not know of a class X until a student registers for it. On the other hand, if we deleted this student from the database, we would not know the school even had class X. This is terrible design for a database and makes the database not very helpful (as it is missing an unknown amount of information) and thus we strive to design complete databases.

2. Consider the sample relation **R(A,B,C,D)** below. Assume the tuples shown completely capture the functional dependencies (FDs) of **R** and that any new tuples would not conflict with them.

A	B	C	D
2	1	2	1
3	2	2	3
1	1	2	1
2	3	2	1

a. [5]Specify a key of **R** that contains the fewest number of attributes. Explain your answer and how you arrived at it.

b. [5] Is there a FD in **R** $\alpha \rightarrow \beta$ where α is not a key? Briefly explain how you arrived at it.

c. [5] Is **R** in BCNF? Briefly explain.

d. [5] Does **R** contain an update anomaly? Briefly explain. But use a specific example.

3. Consider a relation **R(A,B,C,D)** with a set of functional dependencies **F** where
   A → B, B → C, CD → B

   a. [5] What are all of the possible superkeys of **R**? Explain.

   b. [5] Give an additional completely non-trivial functional dependency that is in $\mathbf{F}^+$ (but not in **F**) and explain how you arrived at it.

   c. [10] Put **R** in BCNF clearly stating at each step the FDs and the keys of the decomposed relations.

Corrections - CS345 Exam 2 Fall 2023 **Name Jax Lubkowitz**

4. Consider the relation **`gamer(email, name, games, consoles).`** A person who plays video games has one email address, a name, plays several different video games, and has several gaming consoles. For example the gamer, goober@goober.com has a name Poindexter, and likes to play ***The Legend of Smellda***, ***Super Mango,*** and has a ***PlayStation, XBox, and a Nintendo Switch***.

   a. [5] For goober@goober.com how many rows in **gamer** would we need to insert in order to accurately capture all of the information above? Explain your answer.

   b. [5] What are all of the the completely non-trivial functional dependencies in the **gamer** relation? Explain answer.
   Email → Name
   Email, Games → Name
   Email, Consoles → Name
   Email, Games, Consoles → Name

   In the Gamer database, each email distinguishes one name and can distinguish multiple games and consoles. This means that Email determines name and multiplicatively determines Consoles and Games (not part of FDs). Then by augmentation, we can add Games, Consoles to Email to also determine Name as each row in Gamer has a unique email and any combination of Consoles/Games with that email will still determine the proper name. There are no *non-trivial* functional dependencies that determines Games and Consoles as in the current gamer relation there is multiple rows with the same email and name corresponding to differing Games and Consoles.

   c. [5] Using only the functional dependencies from part b, is **gamer** in BCNF? If not, put it in BCNF.
   No, it is not in BCNF as Gamer has multiple non-trivial functional dependencies.

   1) Gamer(Email, Name, Games, Consoles)
   Email → Name
   Email, Games → Name
   Email, Consoles → Name
   Email, Games, Consoles → Name
   2.) r1 = (Email, Name) **r2 = (Email, Games, Consoles)**
   Email → Name **Email, Games, Consoles → Email, Games, Consoles**
   Key = Email **Key = All-Key**
   Now in BCNF

d. [5] Does gamer contain any multivalue dependencies? If so, what are they? Briefly explain.

e. [5] Is **gamer** in 4th NF? If not, put it in 4th NF.

5. [10] Consider a relation **R(A,B,C,D)** that has a massive number of rows in it and we need to determine if the functional dependency **BC → D** was consistent with the rows of **R**. Write an SQL query that evaluates to a scalar of the number of rows that violate the functional dependency. Explain your answer. On Ada and in Postgres create a database named ***yourusername*_exam2** add a relation R with the four attributes, add lots of dummy test data (use AI), and analyze the test data to make sure that when you run your query you are getting sensible results. Paste your query, your test data, and the output of the query into the space below. Make sure it is formatted neatly.

Corrections - CS345 Exam 2 Fall 2023 **Name Jax Lubkowitz**

6. [10] Draw an **ER** diagram for the Books/Users/Ratings database from **HW5**. Be as complete as possible. Paste an image from Visual Paradigm.