DASC 5333 Database Systems for Data Science
Spring 2025
Section 1 Mid-Term Examination

Last Name: _____ **First Name:** _____ **Student Id:** _____

Number: _____

Time allowed: *1 hour 20 minutes.* Total score: 101 points. *Closed* book examination. *A letter-size information sheet (both sides) prepared by yourself is allowed.*

Answer all questions. <u>Turn in everything: question and answer papers, information sheet and sketch papers. They will be stapled together.</u>

(1) [30 points] The goal is to build a toy prototype application below. Construct an UML class diagram to capture and model the partial requirements. You should list class names, attributes with multiplicities, and associations with multiplicities. The roles of associations should also be provided when appropriate. Multiplicities should be as specific as possible. Show the stereotypes <<pk>> and/or <<unique>> (indicating that the value of the attribute must be unique for each object) when applicable. Since this is only a simplified part of the application, model your design in a flexible way.

Simplified Toy Library System (STLS)

The library system has many branches, each with a name and a unique branch id. Libraries have books. The system stores the ISBN (a unique identifier defined by the Library of Congress), title, edition, publication year, the primary author, and co-authors of a book. Note that information about the edition and the year of publication may not be available. Authors are people in which STLS stores their names. The database stores a unique Id for a person. A book may have no author. A person can be an author or an editor of multiple books. Two people may have the same name.

For example, for our textbook, "Database Illuminated," its ISBN is 978-284-05694 and the publication year is 2017. The primary author is Catherine Ricardo. The co-author is Susan Urban.

The library system may acquire multiple copies of the same book. Each copy has a unique inventory Id. When available, the date of acquisition is recorded. STLS also stores the branch location where a book copy is currently resided in.

Please answer your question on the next page.

(1)   Your answer here:

(2) [15 points] Consider the following data model in the UML class diagram. Attribute multiplicities are included. Construct a reasonable set of relation schema to implement it. For each relation, list its candidate keys, foreign keys, and all attributes you know for sure that are nullable and non-nullable. Indicate whether a surrogate primary key is created.
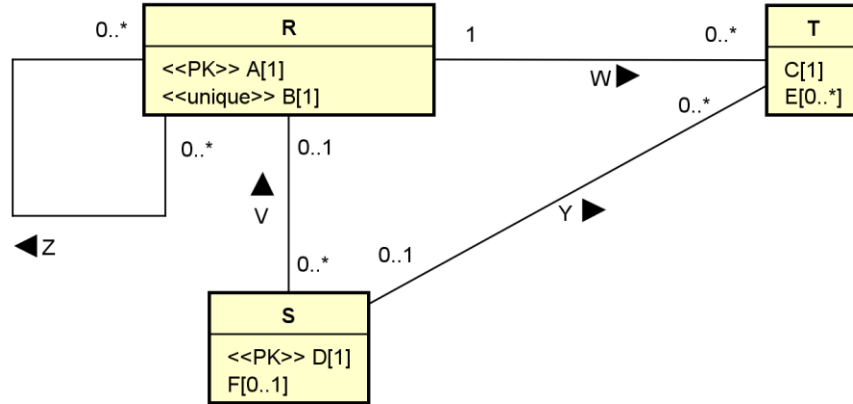
Answer: fill in the table below.

Relation		Relation	
[CK]		[CK]	
[FK]		[FK]	
[Nullable]		[Nullable]	
[Non-nullable]		[Non-nullable]	
[Note]		[Note]	
Relation		**Relation**	
[CK]		[CK]	
[FK]		[FK]	
[Nullable]		[Nullable]	
[Non-nullable]		[Non-nullable]	
[Note]		[Note]	
Relation		**Relation**	
[CK]		[CK]	
[FK]		[FK]	
[Nullable]		[Nullable]	
[Non-nullable]		[Non-nullable]	
[Note]		[Note]	

(3) [26 points] True or False. *Circle* the choice or write 'T' or 'F' *clearly*.

(a) [T or F] SQL is based on an object-oriented model.

(b) [T or F] A SQL SELECT statement may have no FROM clause.

(c) [T or F] In the relational model, a relation instance is a set of attribute names.

(d) [T or F] An advantage of DBMS as compared to file systems is its higher capability for providing secure access control.

(e) [T or F] It is possible for the relation R(A,B,C,D) to have no candidate key.

(f) [T or F] In SQL, the execution of an UPDATE statement can only update one row in a table at a time.

(g) [T or F] The class diagram in UML is a kind of behavioral diagram.

(h) [T or F] Knowing that {A, B} is the primary key of R(A,B,C,D), {A,C} cannot be a candidate key.

(i) [T or F] The data model of Excel spreadsheet is based on set theory.

(j) [T or F] In SQL, null values represent missing or inapplicable information.

(k) [T or F] GPA can be a reasonable example of a derived attribute in databases.

(l) [T or F] If R(A, B) has six rows and S(A,C) has four rows, the SQL statement "SELECT * FROM R INNER JOIN S ON (R.A = S.A);" may return 0 to 24 rows.

(m) [T or F] A foreign key may reference any column in the parent table.

Question 4 uses the toyu database, which is provided separately.

(4) [30 points] Write the *SQL* queries for the following data problems. Result orders are unimportant unless explicitly stated otherwise.

(a) Show the names of all students minoring in CINF, and the classId and grades of their enrolled classes in the following manner.

```
+-----------+---------+-------+
| student   | classId | grade |
+-----------+---------+-------+
| Tony Hawk |   10000 | A     |
| Tony Hawk |   10001 | A     |
| Tony Hawk |   10002 | B+    |
| Tony Hawk |   10003 | C     |
| Tony Hawk |   10004 | A-    |
| Tony Hawk |   11001 | D     |
| Mary Hawk |   10000 | NULL  |
| Mary Hawk |   10001 | A-    |
+-----------+---------+-------+
8 rows in set
```

(b) Show the ids and names of students who have a grade of B+ in one enrolled class and a grade of D in another enrolled class.

```
+--------+------------+
| stuId  | student    |
+--------+------------+
| 100002 | David Hawk |
| 100000 | Tony Hawk  |
+--------+------------+
2 rows in set
```

(c) Show the names and majors (full names) of students who have taken a class from a CINF faculty member in the following manner.

```
+----------------+----------------------------+
| student        | major                      |
+----------------+----------------------------+
| Linda Johnson  | Computer Information Systems |
| Lillian Johnson | Computer Information Systems |
+----------------+----------------------------+
2 rows in set
```