Midterm 15 March 2019, questions and answers

Database Systems (City University of Hong Kong)

Problem One: Basic Concepts (27 points)
1)Judge whether each of the following statement is TRUE or FALSE and give your justification for your answer. [15 points]

//Marking tips: Correct answer is worth 2 points; correct explanation 1 point!

(a) A relational database schema consists of all the data of the database at particular moment.
***False**: A "schema" consists of a set of table definitions. OR A database state/instance consists of of all the data of the database at particular moment.*

(b) All the relationships involving a weak entity set should connect to that weak entity set using double line
False. Double line involving a weak entity set means identifying relationship. Only the entities from which this weak entity is borrowing primary keys need to connect to the weak entity using double line.

(c) The SQL statement SELECT * FROM table WHERE a < 10 or a >= 10 is equivalent to SELECT * FROM table WHERE true.
False. *The first one omits null values of a.*

(d) If a 3NF table has only one candidate key, then this table is also in BCNF.
.
True. *BCNF is different from 3NF only when the table has two or more candidate keys which are overlapping.*

(e) "Update" on a view V is safe and allowable only if V is defined/derived from no more than two base tables/relations.
***False**: A view V can be updated only if V is defined/derived from ONE base table.*

2) Answer the following questions using not more than five (5) sentences for each:
(a) Explain the difference between the not null constraint and primary key constraint. Which one is more strict? [6 points]

Answer: Primary key constraint is more strick [3 point]
Both primary key constraint and not null constraint do not allow null. Besides, primary key constraint requires unique value.[3 points]

(b) Explain the difference between a data definition language (DDL) and a data manipulation language (DML). List SQL commands respectively belongs to DDL and DML. [6 points]

Answer: The data definition language is used to describe external and logical schemas. [1 point]
The data manipulation language is used to access and update data. [1point]
For SQL
* DDL: Create Table [1 point] * DML: Insert, Delete, Update[3 points]

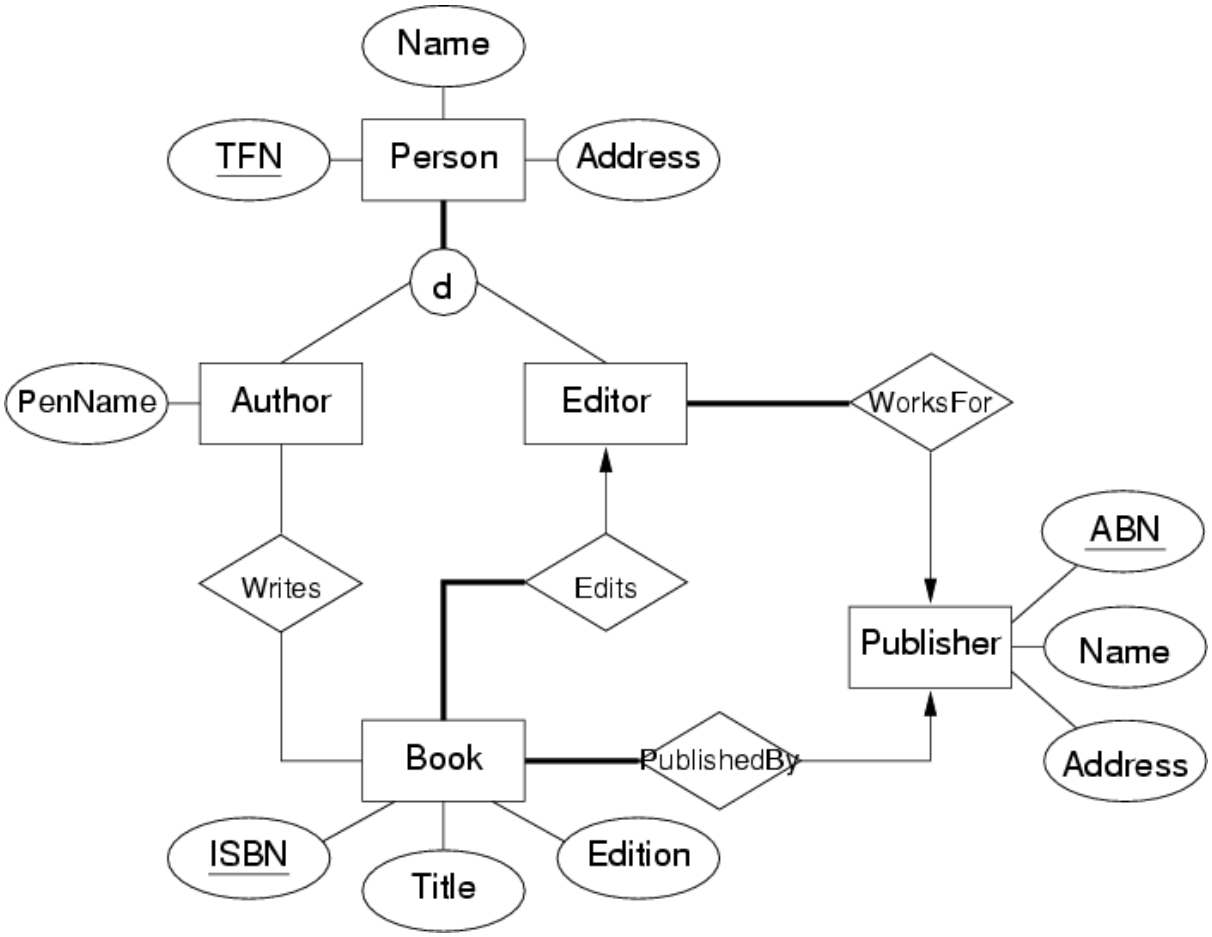
Problem Two: ER model and Relational Model (27 points)

(a) Consider the following set of requirements to model some aspects of the book publishing industry:

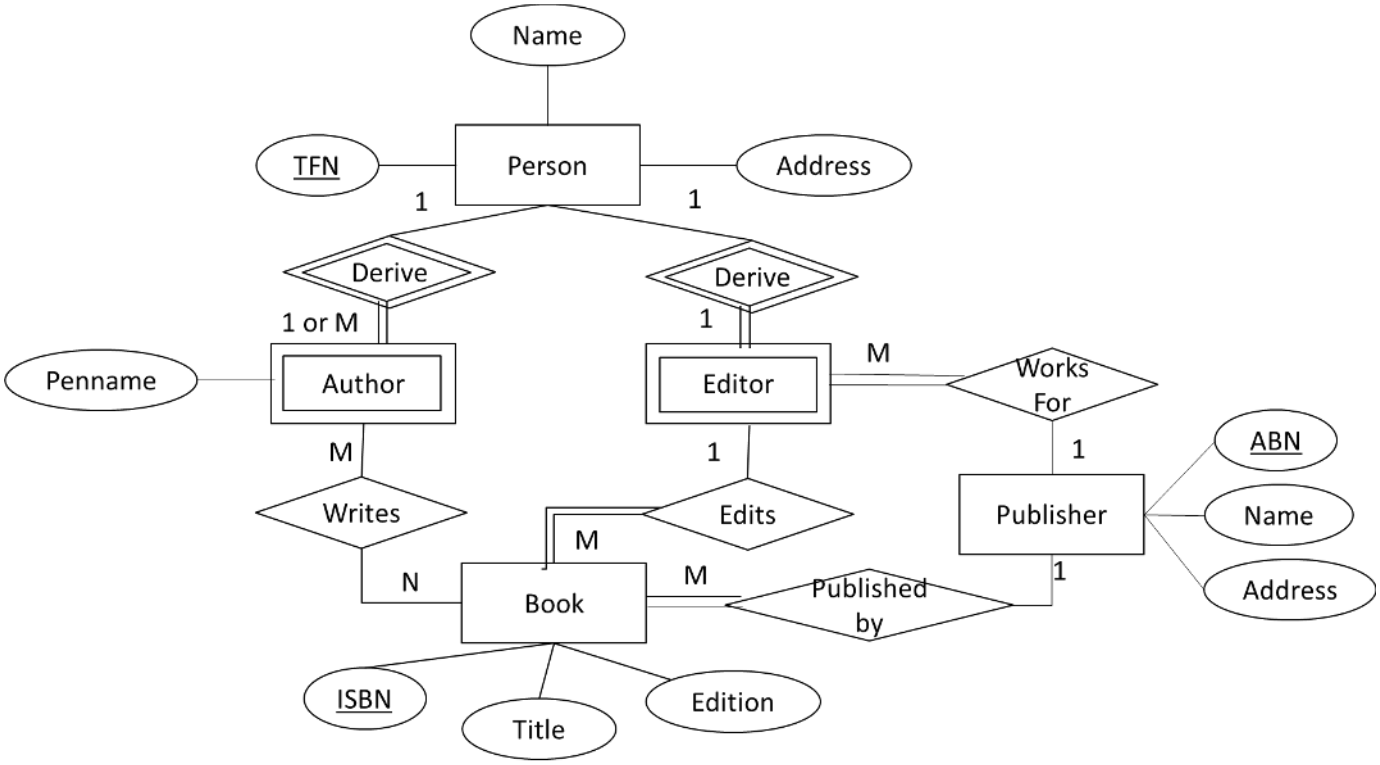
- there are two kinds of people to consider: authors and editors
- for each person, we need to record their tax file number (TFN), their real name, and their address • Each person is identified by a distinct tax file number
- authors write books, and may publish books using a "pen-name" (a name which appears as the author of the book and is different to their real name)
- editors ensure that books are written in a manner that is suitable for publication
- every editor works for just one publisher
- editors and authors have quite different skills; someone who is an editor cannot be an author, and vice versa *//can ignore this requirement. this is not talked in the lecture*
- a book may have several authors, just one author, or no authors (published anonymously)
- every book has one editor assigned to it, who liaises with the author(s) in getting the book ready for publication
- each book has a title, and an edition number (e.g. 1st, 2nd, 3rd)
- each published book is assigned a unique 13-digit number (its ISBN); different editions of the same book will have different ISBNs
- publishers are companies that publish (market/distribute) books
- each publisher is required to have a unique business number (BN)
- a publisher also has a name and address that need to be recorded
- a particular edition of a book is published by exactly one publisher

Design an ER diagram for the above application. Note that you must: a) underline all primary key attributes b) clearly indicate relationship cardinalities c) clearly indicate participation constraints d) choose sensible names for attributes, entities and relationships e) use a single, well-recognized ER notation for the entire diagram . [16 points]

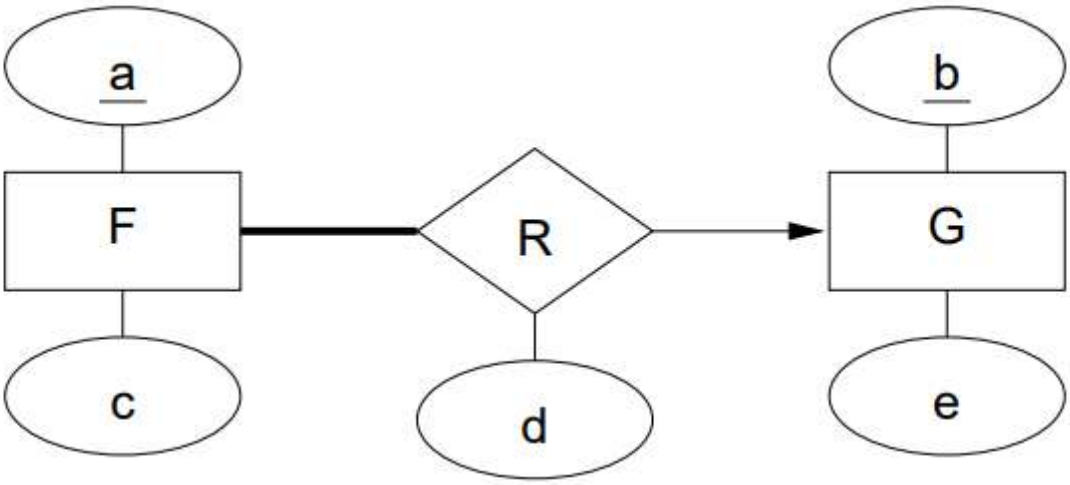
// Each entity set counts for 1 points (total 5points); attribute sets for each entity counts for 1 points (total 5points); Each relationship counts for 1 points(total 6 points).
//This is just sample answers. The student can earn the credits if their diagram is different but correct.



OR



(b) Convert the ER design below into a relational schema, expressed as a collection of SQL **CREATE TABLE** statements.



Assume that the primary keys are integers, and all other attributes are long strings (use type varchar(256)). Your schema must show all primary key constraints, foreign key constraints and "not null" constraints suggested by the diagram. [11 points]

Answer:

// Each line worth 1 points total 11 points

```
create table F (
      a      integer,
      c      varchar(256),
      R      integer not null, or interger,
      d      varchar(256),
      primary key (a),
      foreign key (R) references G(b)
   );
   create table G (
      b      integer,
      e      varchar(256),
      primary key (b)
   );
```

//It is also correct if the student creates an additional table for the relationship R.

Problem Three: SQL and Relational Algebra (24 points)
Consider the following relations containing airline flight information:

Flights(flno: integer, from: string, to: string, distance: integer, departs: time, arrives: time)
Aircraft(aid: integer, aname: string, cruisingrange: integer)
Certified(eid: integer, aid: integer)
Employees(eid: integer, ename: string, salary: integer, nationality:string)

Note that the Employees relation describes pilots and other kinds of employees as well; every pilot is certified for some aircraft (otherwise, he or she would not qualify as a pilot), and only pilots are certified to fly.

Write the following queries in BOTH relational algebra and SQL.
// Each relational algebra expression counts for 4 points and SQL statements counts for 4 points

1. Find the all the aids of aircrafts for which French pilots are certified. [8 points]

$\pi_{aid}(\sigma_{nationality='French'}(Employees \bowtie Certified))$

```
SELECT C.aid
FROM Employees E, Certified C
WHERE E.eid = C.eid AND E.nationality = 'French'
```

2. Find the names of pilots certified for some Airbus aircraft. [8 points]

$\Pi_{name}(\sigma_{aname='Airbus'}(Aircraft \bowtie Certified \bowtie Employees))$

```
SELECT C.name
FROM Aircraft A, Certified C, Employees E
WHERE A.aid = C.aid AND A.aname = 'Airbus' AND E.eid = C.eid
```

3. Identify the flights that can be piloted by at least one pilot whose salary is NO more than $100,000. (Hints: The aircraft can only be used in the flight when the flight distance is smaller than the cruising range)

$\pi_{flno}(\sigma_{distance< cruisingrange\ and\ salara <= 100,000}(Flight \bowtie Aircraft \bowtie Certified \bowtie Employees))$

```
SELECT F.flno
FROM Aircraft A, Certified C, Employees E, Flights F
WHERE A.aid = C.aid AND E.eid = C.eid AND
        distance < cruisingrange AND salary <= 100,000
```

Problem Four: Constrains and Normalization [22 points]

Suppose we have a relational database table R for the published books, consisting of the following attributes: I(ISBN number), T(Book_title), N(Author_name), Y(Book_type), M(List_price), A(Author_affil), P(Publisher) with the following functional dependencies:{I → T, I → N, T → Y, Y → M, N → A, T → P, TN → I}.

(a) Find all the candidate keys for the relation table R. [6 points]

Answer: R has two cand. keys: {I} and {T,N}

//Finding one candidate key is worth 3 points. Finding both candidate keys can get full 6 points. If one answers more than two candidate keys, then, if the correct ones are also contained, s/he gets 3 points, otherwise, 0.

(b) Determine the highest normal form the following decomposition is in. Justify your answer with a brief explanation. [8 points]

R1(T, Y, M, P)

R2(N, A)

R3(I, T, N)

Answer: R1 in 2NF, R2 and R3 in BCNF, overall in 2NF.

//Judging a table correctly is worth roughly 2 points, overall judgement of the set of tables is worth 2 point. Nearly correct judgement gets partial points.

(c) Come up with a decomposition for R so that the tables are in Boyce-Codd normal form (BCNF). [8 points]

Answer:

R1(T, Y, P)

R2(N, A)

R3(I, T, N)

R4 (Y,M)

Based on the answer of part (b), R2 and R3 are already in BCNF, so we only need to split R1 into two tables: (T,Y,P) and (Y,M)

//Each correct table in BCNF is worth 2 point.