CISC437/637 Database Systems Midterm Exam

Above please write your name and whether you are in 437 or 637.

You have from 10:10 to 11:00am to complete the following questions. No notes, no book, no wireless (or wired) devices are allowed. Use the back of the page if you need more space. Good luck!

Multiple Choice [36 points]

1. A "superkey" is:
   (a) an attribute or set of attributes that could be the primary key
   (b) functionally dependent on non-key attributes
   (c) any set of fields that determines the values of all other fields
   (d) the primary key of a relation

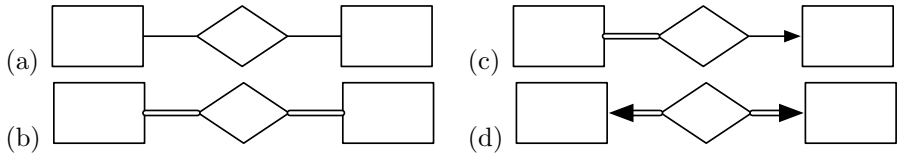
2. When a record in relation R with primary key K is deleted, records in relation R' with foreign key $R.K$ are also deleted. This is called a:
   (a) incremental delete
   (b) incremental update
   (c) cascading delete
   (d) cascading update

3. Which built-in SQL function computes the average of values in a field?
   (a) AVG
   (b) MAX
   (c) COUNT
   (d) SUM
   (e) MIN

4. Which set of SQL keywords is used for data definition in an existing table?
   (a) CREATE, ALTER, DROP
   (b) INSERT, UPDATE, DELETE
   (c) SELECT, FROM, WHERE
   (d) GROUP BY, ORDER BY, HAVING

5. In a relational database management system, a rule that ensures that every record in a table is unique is called a ...
   (a) mapping constraint
   (b) referential integrity constraint
   (c) key constraint
   (d) participation constraint

6. "Each flight arrives at exactly one airport." Which ER diagram best captures this requirement?



7. Which of the following is *not* required for a relation to be in 1NF?

   (a) All values for a field must come from the same domain.
   (b) The values of fields must be atomic (not lists or sets).
   (c) No two rows may be identical.
   (d) The columns must be in a particular order.

8. Which pair of SQL keywords is used to create and delete records in a table?

   (a) CREATE, DROP
   (b) CREATE, DELETE
   (c) CREATE, ALTER
   (d) INSERT, DROP
   (e) INSERT, DELETE
   (f) INSERT, ALTER

9. Which of the following is always true?

   (a) Relations in 1NF may have list values in record fields.
   (b) Every relation that's in BCNF is also in 3NF.
   (c) Every relation that's in 3NF is also in BCNF.
   (d) Every relation that's in 1NF is also in BCNF.

10. Which pair of SQL keywords is used to create and delete a table?

    (a) INSERT, DROP
    (b) INSERT, DELETE
    (c) INSERT, ALTER
    (d) CREATE, DROP
    (e) CREATE, DELETE
    (f) CREATE, ALTER

11. Which of the following is used in an E-R diagram to represent a one-to-one relationship between two entity sets, in which every entity in both sets participates?

    (a) A double-headed double-line arrow between the entity sets.
    (b) Single-line arrows from the relationship set to each entity set.
    (c) Double-line arrows from the relationship set to each entity set.
    (d) Double lines with no arrow from the relationship set to each entity set.
    (e) A single-line arrow from the relationship set to one entity set, and a double-line arrow from the relationship set to the other entity set.

The next seven questions are based on the following tables:

airline_code	name	hub
TAA	Trans American Airlines	PHL
PNA	Puño Airlines	BCN
OA	Oceanic Airlines	LAX
PA	Paradise Airlines	PHL
QA	Quantum Airlines	DUB

(a) Airlines(airline_code, name, hub)

airport_code	name	location
PHL	Philadelphia Int'l Airport	Philadelphia, PA
PIT	Pittsburgh Int'l Airport	Pittsburgh, PA
BCN	Aeropuerto de Barcelona	Barcelona, Spain
DUB	Aerfort Bhaile Átha Cliath	Dublin, Ireland
LAX	Los Angeles Int'l Airport	Los Angeles, CA

(b) Airports(airport_code, name, location)

airline_code	flightNo	origin	destination	depart	arrive	price
TAA	1010	PHL	LAX	10:00	12:30	$250
OA	815	SYD	LAX	13:00	09:30	$600
TAA	1020	LAX	PHL	14:00	22:30	$250
PNA	1010	PHL	BCN	18:00	08:00	$350
PA	955	PHL	PIT	21:00	22:15	$80
QA	2200	PHL	DUB	21:00	08:30	$300
TAA	1988	PHL	PIT	22:00	23:15	$75

(c) Flights(airline_code, flightNo, origin, destination, depart, arrive, price)

12. To find the origin and destination cities of scheduled flights, which SQL statement would you use?

    (a) SELECT location FROM Flights, Airports
    WHERE origin = airport_code

    (b) SELECT location FROM Flights, Airports
    WHERE destination = airport_code

    (c) SELECT A1.location, A2.location FROM Flights, Airports A1, Airports A2
    WHERE origin = A1.airport_code AND destination = A2.airport_code

    (d) SELECT location FROM Flights, Airports
    WHERE origin = airport_code AND destination = airport_code

13. A right outer join of Airports and Flights on "airport_code = origin" (Airports $\bowtie\!\!\!\!\subset_{airport\_code=origin}$ Flights) will return how many records?

    (a) 4

    (b) 5

    (c) 6

    (d) 7

14. Joining one instance of Flights to another with the condition that destination in the first equals origin in the second will return how many records?

    (a) 5

    (b) 6

    (c) 7

    (d) 8

    (e) 49

15. Which of the following would find the flights operated by airlines starting with the letter 'P'?

    (a) SELECT F.* FROM Airlines A, Flights F
    WHERE A.airline_code=F.airline_code AND A.name='P*'

    (b) SELECT F.* FROM Airlines A, Flights F
    WHERE A.airline_code=F.airline_code AND A.name LIKE 'P%'

    (c) SELECT F.* FROM Airlines A, Flights F
    WHERE A.airline_code=F.airline_code AND A.name='P%'

    (d) SELECT F.* FROM Airlines A, Flights F
    WHERE A.airline_code=F.airport_code AND A.name LIKE 'P%'

16. Which of the following would find the average price of flights by airline?

    (a) SELECT airline_code, AVERAGE(price) FROM Flights GROUP BY price

    (b) SELECT AVERAGE(price) FROM Flights

    (c) SELECT airline_code, AVERAGE(price) FROM Flights GROUP BY airline_code

    (d) SELECT airline_code, AVERAGE(price) FROM Flights

17. Suppose Flights is defined in SQL with the referential integrity constraint FOREIGN KEY (airline_code) REFERENCES Airlines(airline_code) ON DELETE CASCADE. What would happen if TAA was deleted from Airlines?

    (a) All TAA flights would be deleted from Flights.

    (b) All TAA flights would be reassigned to a default operator.

    (c) All TAA flights would be given a null value for airline.

    (d) Nothing.

18. Which of the following commands will fail given the schema and the data in the tables?

    (a) INSERT INTO Flights VALUES ('TAA', 1010, 'PHL', 'BCN', '18:00', '08:00', 350)

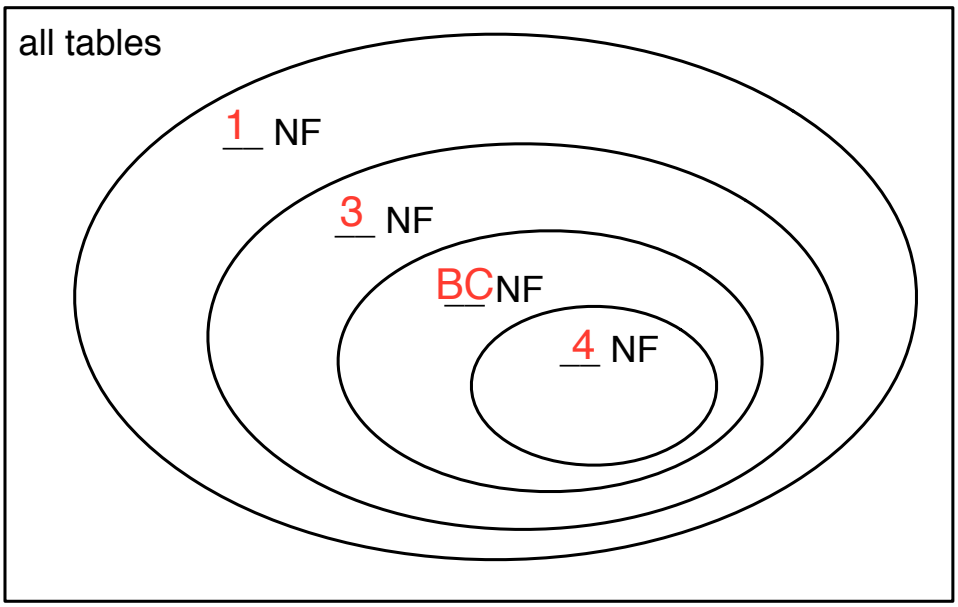
    (b) INSERT INTO Flights VALUES ('QA', 1010, 'PHL', 'BCN', '18:00', '08:00', 350)

    (c) INSERT INTO Flights (airline_code, flightNo, origin, destination, depart, arrive, price)
    VALUES ('QA', 1010, 'PHL', 'BCN', '18:00', '08:00', 350)

    (d) INSERT INTO Flights (airline_code, flightNo) VALUES ('QA', 1010)

Short Answer [32 points]

1. The Venn diagram below represents the universe of all possible tables. Each circle represents the set of relations in one of the four normal forms we talked about in class (in random order: 3NF, 1NF, 4NF, BCNF) from least strict to most strict. Fill in the blanks to complete the name of the normal form each circle represents.



2. Consider the following three relational schema:

   Airlines(airline_code, name, hub)
   Airports(airport_code, name, location)
   Flights(airline_code, flightNo, origin, destination, depart, arrive)

   Express each of the provided information needs as a relational algebra query using only projection (π), selection (σ), renaming (ρ), inner joins ($\bowtie$), and set union/intersection ($\cup$, $\cap$).

   (a) Find the origin and destination of flights that depart after 5pm.

   (b) Find airlines that can get from Los Angeles to Pittsburgh with exactly one stopover. (For example, in the relation instance on page 3, Trans American Airlines has a flight from LAX to PHL and PHL to PIT. It would qualify.)

airline_code	name	hub	hubName	location
TAA	Trans American Airlines	PHL	Philadelphia International Airport	Philadelphia, PA
PNA	Puño Airlines	BCN	Aeropuerto de Barcelona	Barcelona, Spain
OA	Oceanic Airlines	LAX	Los Angeles International Airport	Los Angeles, CA
PA	Paradise Airlines	PHL	Philadelphia International Airport	Philadelphia, PA
QA	Quantum Airlines	DUB	Aerfort Bhaile Átha Cliath	Dublin, Ireland

3. The relation above contains redundant information.

   (a) Give two examples of problems that redundancy can cause in the database.

   (b) What could you do to eliminate redundancy? Describe specifically what new relations you could create to replace this one.

4. I have created two SQL tables as follows:

```
CREATE TABLE Airports (airport_code CHAR(3),
                       name CHAR(50),
                       location CHAR(30),
                       PRIMARY KEY (code))

CREATE TABLE Airlines (airline_code CHAR(3),
                       name CHAR(30),
                       hub CHAR(3) NOT NULL UNIQUE,
                       PRIMARY KEY (code),
                       FOREIGN KEY (hub) REFERENCES Airports(code)
                         ON UPDATE CASCADE ON DELETE NO ACTION)
```

   Indicate whether each of the following statements is true or false (assuming referential integrity is enforced by the DBMS):

   (a) An airline may have a NULL value for hub.

   (b) Deleting an airport may directly result in airline records being deleted.

   (c) Changing an airport's code may directly result in airline records being updated.

   (d) An airport may be a hub for two or more airlines.

   (e) An airline may have two or more hubs.

   (f) Every airline must have a hub.

Database Design [32 points]

These four questions are all related to the following requirements:

(a) Airlines and airports are identified by a short code assigned by the Federal Aviation Administration (FAA). No two airlines or airports have the same code.

(b) Airports are located in a particular city (in exactly one city).

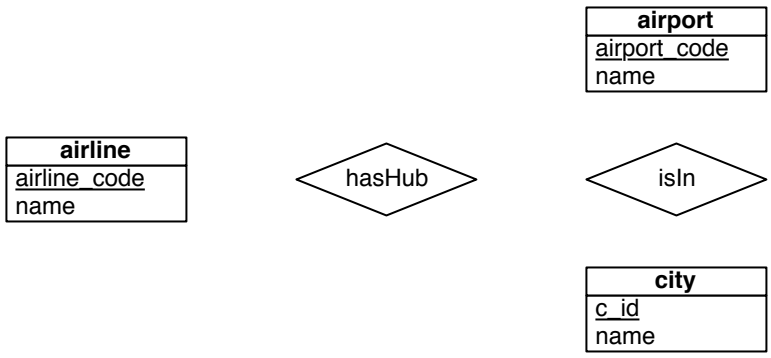
(c) Each airline has *one or more* hub airports that serve as major transfer points for that airline.

(d) An airline can have at most one hub in any city.

(e) Some airports are not hubs for any airline; some are hubs for more than one airline.

Answer the following questions:

1. Entity sets and relationship sets are shown below. Draw in participation lines and mapping constraint arrows between them to capture the requirements above as closely as possible.



2. I have created the following five relations to try to capture the requirements above:

   Airport(airport_code, name)
   City(c_id, name)
   InCity(airport_code, c_id)

   Airline(airline_code, name)
   HasHub(airline_code, airport_code)

   (a) Underline the fields that should be chosen as primary keys for each relation.

   (b) Which (if any) of the requirements above are *not* captured by this set of relations? (You may just write the list of letters corresponding to requirements.)

   (c) Which two relations could be combined into one without losing any power to capture requirements? Write out the combined relation's schema.

3. Now I drop the HasHub relation and create a new relation called HasHubInCity:

   HasHubInCity(airline_code, c_id, airport_code)

   (a) This relation captures the functional dependency `airline_code, c_id` → `airport_code`. Which requirement does this FD correspond to?

   (b) Suppose HasHubInCity has been populated with the following data:

airline_code	c_id	airport_code
TAA	001	PHL
PNA	002	BCN
OA	003	LAX
PA	001	PHL
QA	004	DUB

   (City codes are defined as follows: 001 is Philadelpha, 002 is Barcelona, 003 is Los Angeles, and 004 is Dublin.)

   Give an example of a new row that is *valid* according to the relational schema but *invalid* in reality.

4. I have proposed two possible solutions for modeling hubs: either HasHub(airline_code, airport_code) or HasHubInCity(airline_code, c_id, airport_code). We must decide on one of them.

   (a) Is HasHub in Boyce-Codd Normal Form (BCNF)? Is HasHubInCity? Explain why or why not.

   (b) Is HasHub in Third Normal Form (3NF)? Is HasHubInCity? Explain why or why not.

   (c) For extra credit, provide a solution that is not in 1NF.