Midterm Exam #3 DRAFT SOLUTION

Question 1 - Databases are Fun! (8 points)

Name the type of schema (physical, conceptual, or external) that is considered in each of the following scenarios:

A. The schema a user sees on Fakebook when editing their profile. (1 point)

External schema

B. The binary files stored in Fakebook's private server that contains all users' personal info. (1 point)

Physical Schema

C. The schema you see when creating a new table in the Fakebook SQL database. (1 point)

Conceptual Schema

DishID	DishName	Price AvgRating
21341	Krabby Patty	2.99 10
34563585	Spinach Puffs	5.50 7
23	Ratatouille	27.40 8
3287	Butterbeer	1.99 7
123534	Lembas Bread	3.7 3

Using the above *instances* (i.e. snapshot/subset) of the database of a restaurant in a theme park, answer questions D and E.

D. List all single attributes that are **definitely not** candidate keys. Why are these attributes definitely not candidate keys? If all attributes could be candidate keys, explain why. (2 points)

AvgRating since there are duplicate values

E. Which single attribute is **definitely** a candidate key? Why? If there are no definite candidate keys, explain why. (2 points)

None, we can not say anything is a candidate key based on a snapshot

F. True or False, you can ALWAYS decompose any relation into a set of relations that are all in 3NF or all in BCNF? (1 point)

True

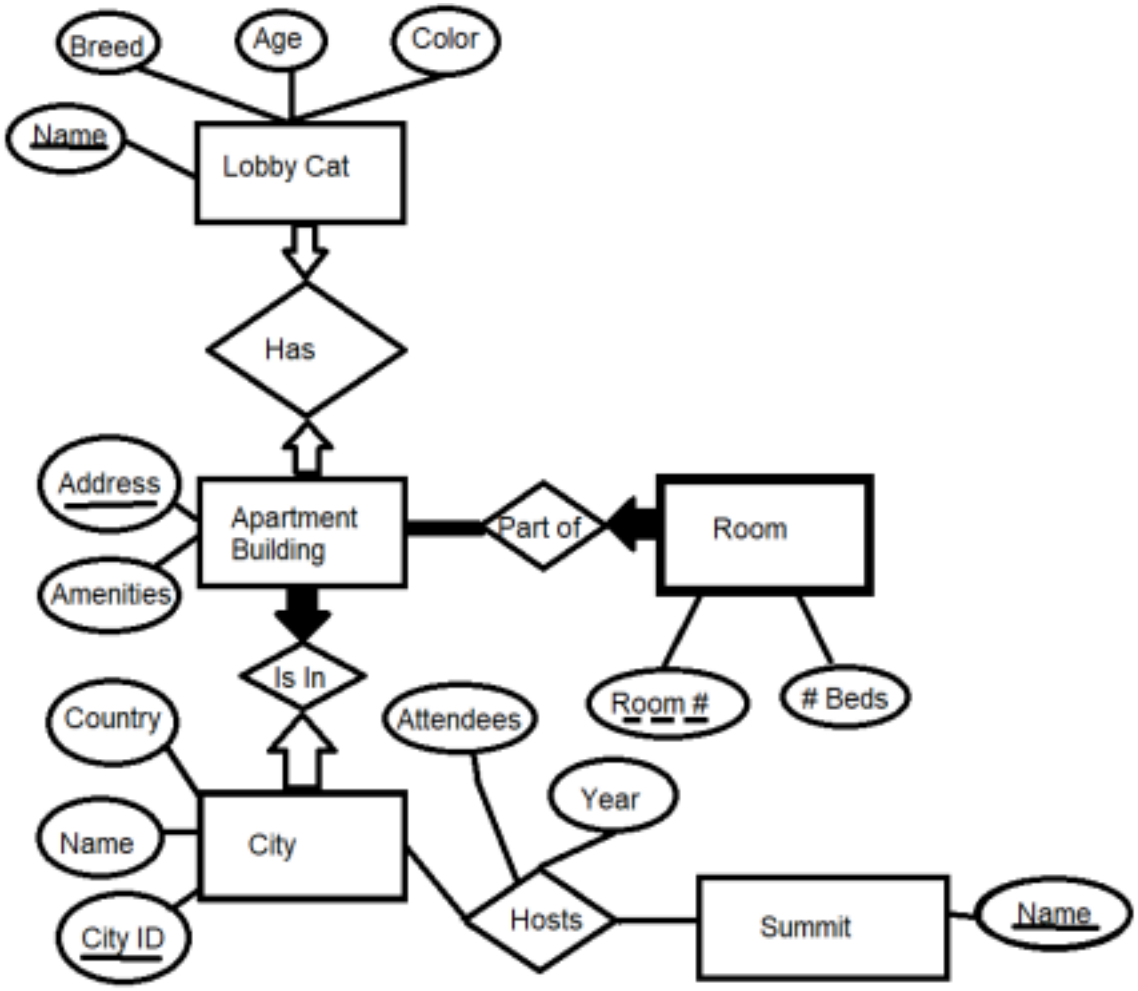
Question 2 - Sahil Towers (22 points)

Thanks to its recent success, Sahil Towers™ is planning on expanding worldwide to bring college students affordable and aesthetic housing. To help with this, you've been asked to develop a database for the company.

Develop an ER diagram the reflects the constraints below:

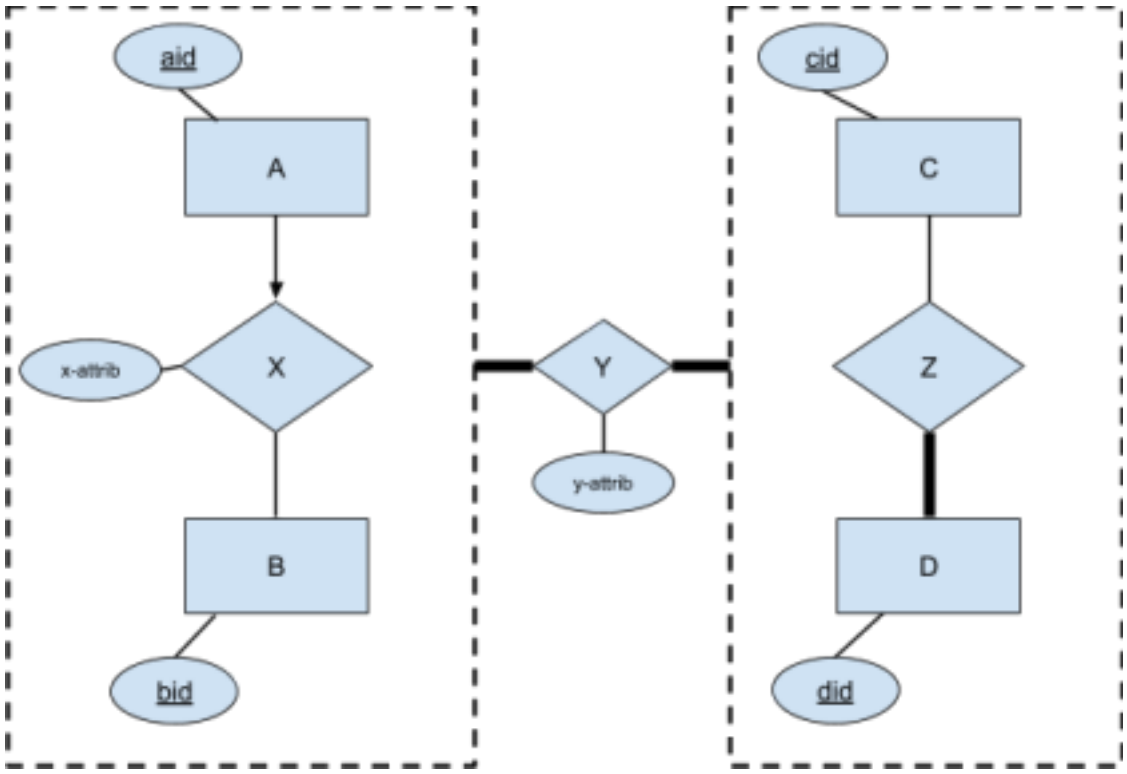
- Each apartment building is uniquely identified by its address. We also keep track of the number of amenities of each apartment building.
- An apartment building may have at most one cat that lives in its lobby ○ Each cat has a unique name. They also have a breed, an age, and a colour. ○ Each apartment building can have up to one cat and each cat can live in a single apartment building. Unfortunately, not every cat is in a Sahil Apartment Building
- Each apartment building has at least one room in it
  - ○ These rooms are uniquely identified by a combination of the room number and the apartment building's address. Each room has a number of beds.
  - ○ A room cannot exist without the single apartment building it is located in
- Each apartment building must be located in exactly one city
  - ○ The city has a unique city ID tracked by the Sahil Towers system, a name, and a country it is located in.
  - ○ Each city may have at most one apartment building
- Some cities host summits
  - ○ Each summit has a unique name that does not change from year to year. ○ When a city hosts a summit, we keep track of which year it was hosted in and how many people attended the summit. A city may not hold the same summit multiple times.

You may draw your ER diagram in the space below or draw it out on a separate piece of paper and submit it with your exam.



Question 3 - ER Diagrams are Abstract Art (5 points)

Determine whether the following statements about the given ER Diagram are true or false



A. The set of attributes (cid, did) is a candidate key for relationship Z. (1 point)

True

B. Because every instance of relationship Z participates in Y, and records in C participate in relationship Z, every record in C participates in relationship Y. (1 point)

False

C. Two instances of relationship X ***can't*** have the same value for x-attrib. (1 point)

False

D. aid is a candidate key for the relationship X. (1 point)

True

E. The set of attributes (aid, bid, cid, did) is a candidate key for relationship Y. (1 point)

False

Question 4 - A Night at the Movies (22 points)

Consider a database consisting of the following tables. Primary keys are underlined and attributes with the same name are foreign keys. Assume that none of the attributes are NULLs.

Movies(MovieID, MovieName, MovieYear)
Actors(ActorID, FirstName, LastName)
Castings(MovieID, ActorID)
Ratings(ReviewerID, MovieID, Rating)
Reviewers(ReviewerID, UserName)

A. Write a SQL query that finds all movie names of movies with at least 1 rating greater than or equal to 7.0 and are released after the year 2013. Results should not include duplicates and should be ordered by MovieName in ascending order. (4 points)

```
SELECT DISTINCT M.MovieName
FROM Movies M
JOIN Ratings R ON M.MovieID = R.MovieID AND R.Rating >=
7.0 WHERE M.MovieYear > 2013
ORDER BY M.MovieName ASC;
```

B. Write a SQL query that finds all the MovieIDs of the movies that have more than 5 actors acting in it. (3 points)

```
SELECT DISTINCT MovieID
FROM Castings
GROUP BY MovieID
HAVING COUNT(*) > 5;
```

C. Write a SQL query that finds all the UserNames of reviewers that have given at least 10 movies a rating of 6.3 or above. (5 points)

```
SELECT UserName FROM Reviewers WHERE ReviewerID IN (
SELECT Re.ReviewerID
FROM Reviewers Re
JOIN Ratings Ra ON Re.ReviewerID = Ra.ReviewerID AND Ra.Rating >=
6.3 GROUP BY Re.ReviewerID
HAVING COUNT(*) >= 10);
```

D. Write a SQL query that creates a view called FanFavorites. In the view, store the IDs and average ratings of each movie that has more than 1000 ratings and an average rating of 8.0 or above. Alias the column for average rating with ‘AvgRating’. (4 points)

```
CREATE VIEW FanFavorites AS
SELECT MovieID, AVG(R.Rating) AS AvgRating
FROM Ratings R
GROUP BY R.MovieID
HAVING COUNT(*) > 1000 AND AVG(R.RATING) >= 8.0;
```

E. Find the ID of movies that has reviews written by all reviewers. (6 points)

```
SELECT M.MovieID
FROM Movies M
WHERE NOT EXISTS (
        SELECT ReviewerID
        FROM Reviewers
        MINUS
        SELECT R.ReviewerID
        FROM Ratings R
        WHERE R.MovieID = M.MovieID
);
```

Alternate solution:
Selecting from ratings, grouping by movieid, having count = (select count from reviewers)

Question 5 - The 484 Fruit Stand (18 points)

Use the following schema for this question. Primary keys are underlined and attributes with the same name are foreign keys

StaffMember(Uniqname, FirstName, LastName, Role, CityID)
City(CityID, CityName, StateOrProvince, Country)
Fruit(FruitID, FruitName, VitaminC)
Enjoys(Uniqname, FruitID)

Note: The VitaminC attribute of the Fruit table represents the milligrams of vitamin C that the fruit contains. The CityID field of the StaffMember table refers to the hometown city of staff members. The Enjoys table stores data on which staff members enjoy which fruits.

Given the above schema, write relational **algebra** statements for the following statements: A. Find the first and last names of all staff members whose hometown city is in the country China and who enjoy the fruit whose name is “durian”. (5 points)

$\rho(\text{DurianLovers}, \text{StaffMember} \bowtie_{\text{StaffMember.Uniqname=Enjoys.Uniqname}} \text{Enjoys} \bowtie_{\text{Enjoys.FruitID = Fruit.FruitID}} \sigma_{\text{FruitName='durian'}}(\text{Fruit}))$

$\pi_{\text{FirstName, LastName}}(\text{DurianLovers} \bowtie_{\text{DurianLovers.CityID=City.CityID}}(\sigma_{\text{Country='China'}}(\text{City}))$) Rename operation not necessary

B. Find the name and country of all cities that are NOT a hometown of any staff member whose role is "GSI". (4 points)

$\rho(\text{QualifiedCityIDs}, \pi_{\text{CityID}}(\text{City}) - \pi_{\text{CityID}}(\sigma_{\text{Role='GSI'}}(\text{StaffMember})))$

$\pi_{\text{CityName, Country}}(\text{QualifiedCityIDs} \bowtie_{\text{QualifiedCityIDs.CityID=City.CityID}} \text{City})$

Rename operation not necessary

Note: It is possible for two cities to have the same CityName and Country. If some GSI is from one of those cities, but no GSI is from the other city, it becomes unclear whether or not the CityName and Country pair should be returned in the final result. The instructor solution will include such CityName and Country pairs, but student solutions that don't include it were also accepted.

C. Find the uniqnames of all staff members who enjoy the fruit named 'Mango' or are from the state of 'Ohio'. (4 points)

$\pi_{\text{StaffMember.Uniqname}}(\sigma_{\text{City.StateOrProvince = 'Ohio' v Frut.FruitName = 'Mango'}}(((\text{StaffMembers} \bowtie \text{City}) \bowtie \text{Enjoys}) \bowtie \text{Fruit}))$

D. Find the first names of all the staff members who are not from the state of 'Michigan' and enjoy all the fruits in the database. (5 points)

$\rho(\text{EnjoyAllFruits}, (\pi_{\text{Uniqname, FruitID}}\text{Enjoys}) / (\pi_{\text{FruitID}}\text{Fruit}))$

$\pi_{\text{StaffMember.FirstName}}(\sigma_{\text{City.CityName != 'Michigan'}}\text{EnjoyAllFruits} \bowtie \text{StaffMember} \bowtie \text{City})$ Rename operation not necessary

Question 6 - The Quest for a Normal Table (12 points)

Given a relation R(A,B,C,D,X,Y,Z) with the following set of functional dependencies, circle the option that answers each of the following questions.

$$\{X \rightarrow Y, Y \rightarrow Z, A \rightarrow B, A \rightarrow CD\}$$

A. Use Armstrong's inference axioms (reflexibility, augmentation, transitivity, union, decomposition) to determine which functional dependencies CANNOT be derived. (3 points)

a. $AY \rightarrow AZ$
b. $X \rightarrow Z$
c. $A \rightarrow D$
d. $X \rightarrow A$ Answer

B. Is this relation with columns (A, B, C, D, X, Y, Z) in 3NF form? In BCNF form? (3 points)

a. Both 3NF and BCNF
b. In 3NF but not BCNF
c. In BCNF but not 3NF
d. Neither in 3NF or BCNF Answer

C. If we were to break this relation into two relations, (X, Y, Z) and (A, B, C, D), which of these two relations would be in BCNF form? (3 points)

a. Both (X, Y, Z) and (A, B, C, D)
b. Only (X, Y, Z)
c. Only (A, B, C, D) Answer
d. Neither

D. Which of the following decompositions of the original relation are lossless join decompositions?
(3 points)

a. (A,B,C,D,X) and (X,Y,Z) Answer
b. (A,B,C,D,Y) and (Y,X,Z)
c. (A,B,C,D) and (D,X,Y,Z)
d. Choices A and B
e. Choices A and C

Question 7 - The Fastest Index (3 points)

In a table with T1 with 4 attributes <Tid, Name, Age, Gender>, Tid is the primary key and a clustered index is created on attribute Name. We want to create another index on the table. Circle all of the types of indices that are able to be created. There may be 1 or more

a. An unclustered index on <Tid, Gender> using Alternative 2. Answer
b. A clustered index on <Tid, Name> using Alternative 1.
c. A clustered index on <Age> using Alternative 3.
d. An unclustered index on <Name> using Alternative 3. Answer
e. An unclustered index on <Tid, Age> using Alternative 1.
f. A clustered index on <Gender> using Alternative 2.