Database Systems — CSci 4380
Midterm Exam #1
October 4, 2021

SOLUTIONS

Question 1 (10*3=30 points). You are given the following database. The data model is described in detail in the last page of the exam.

```
stores(storeid, street, state, city, zip)
trucks(license, state, year, mileage, size)
truck_features(license, feature)
renters(username, password, fname, lname)
rentals(rentalid, license, pickup_storeid, dropoff_storeid, username, startdate, enddate, price)
```

(a) Return the `license` plate and `of` size all trucks with a `'low deck' feature` registered in the `'NY' state`, that have zero `mileage` and is due to be picked up on `'10/05/2021'`.

Solutions:

```
R1 = select_{feature = 'low deck'} (TruckFeatures)
R2 = select_{state='NY' and mileage=0} (Trucks)
R3 = select_{startdate = '10/05/2021'} (Rentals)
R4 = project_{license,size} (R1 * R2 * R3)
```

(b) Return the `license` plate of all `'large' size` trucks that are being picked up from a store in a state that is different than the state the truck is registered under.

Solutions:

```
R1 = select_{size='large'} Trucks
R2(l1,s1) = project_{license, state} (Rentals join_{storeid=pickup_storeid}
Stores)
R3 = project_{license} (R2 join_{license=l1 and state<>s1} R1)
```

(c) Return the `first` and `last` name of all renters who **never** had a rental in which the drop off store was in the same state as the pick up store.

Solutions:

```
S1(s1, state1) = project_{storeid, state} (Stores)
R1 = Rentals join_{pickup_storeid=storeid} Stores
R2 = R1 join_{dropoff_storeid=s1 and state1=state} R2
R3 = project_{username} (rentals) - project_{username} (R2)
```

Question 2 (12*3=36 points). Suppose you are given the following relations to add to the data model in the appendix. Answer questions regarding each additional relation below. Please do not use abbreviations' attribute names are shortened to make writing easier.

(a) The truck company also sells many moving supplies like boxes, bags, tape, etc.

`Supplies(item, itemtype, store_id, isavailable, description, price, pickupfee, deliveryfee)`

For each item, there is a specific item type, description, and price.

Whether a specific item is available or not depends on the store as not all stores carry all the items.

An item's pickup and delivery fee depends solely on the item type.

(i) List all functional dependencies that are implied by the above description.

(ii) Based on your functional dependencies, list all the keys and state whether this relation is in BCNF or not. Explain why or why not.

Solutions:

```
item -> itemtype description price
item store_id -> isavailable
itemtype -> pickupfee deliveryfee

Key: item, store_id, itemtype

Not in BCNF as none of the functional dependencies have a superkey on
the left.
```

(b) We store a new relation for booking some help for moving for various tasks, provided by helpers (helperid, name, phone) and booked by users (username) for a specific task.

`MovingHelp(helperid, name, phone, username, task, mdate, street, state, city, zip, rate, hours)`

`helperid` → `name phone`
`username mdate` → `street state city zip hours`
`username task` → `rate`

(i) Can a user book helpers on multiple dates? Answer yes/no and provide a one sentence explanation.

(ii) Can a user pay two different helpers a different price for the same task? Answer yes/no and provide a one sentence explanation.

(iii) What are the keys? Is this relation in 3NF? Please explain why or why not.

(iv) If the relation is not in 3NF, use the 3NF decomposition to find relations that are in 3NF. List each relation and the projected functional dependencies.

Solutions:

(i) Yes, a user can have multiple bookings on different dates

(ii) No, rates are not dependent on helpers, only task.

(iii) Keys: helperid, username, mdate, task. Not in 3NF.

(iv) 3NF decomposition:
(helperid, name, phone), helperid → name phone

(username, mdate, street, state, city, zip, hours), username mdate → street state city zip hours

(username, task, rate) username task → rate

(helperid, username, mdate, task)

(c) We store rental rates for a specific truck size and specific truck features (`ifeat`). Some stores may be excluded from a given rate (`exstore`).

`RentalRates(id, from, to, rate, size, ifeat, exstore)`

This relation has the following functional dependencies:

`id` $\rightarrow$ `from to rate size`
`id` $\Rightarrow$ `ifeat`
`id` $\Rightarrow$ `exstore`

(i) What are the key(s)?
(ii) Is this relation in BCNF? Explain why or why not.
(iii) If it is not in BCNF, use BCNF decomposition to get relations that are in BCNF.
(iv) For all the resulting relations after BCNF decomposition, list if they are in 4NF or not. If not, explain how they can be made into 4NF.

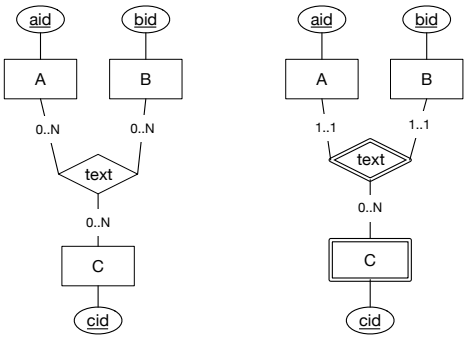
Solutions:

(i) id, ifeat, exstore
(ii) Not in BCNF. The single functional dependency does not have a superkey on the left.
(iii) BCNF decomposition:
RR1 (id, from, to, rate, size), id $\rightarrow$ from to rate size , in BCNF

RR2(id, ifeat, exstore)
(iv) RR1 is in 4NF. No multivalued dependency. RR2 is not in 4NF because the multivalued dependencies do not cover all the attributes.
We must decompose into (id, ifeat) and (id, exstore).

Question 3 (6 points). You are given two alternate models in the following Entity-Relationship diagram. In which ways are these models similar or different? Give a short explanation.

Solutions:



MODEL 1 MODEL 2

In model 1: The id for C is unique in the database and we can store C without any As or Bs. In model 2, C's key is only unique in combination with keys of entities A and B and we need have A's and B's to be able to store C's.

We will also accept: In model 1, we end up storing an extra relation for all different C values when we convert it to the relational data model. In Model 2, we do not store this value.

However, it is not true that for a given C, there is a single A or B in Model 2. When you convert model 2 to relational model, you will get: C(aid, bid, cid). Hence, for a given cid, you can have many As and Bs. The relationship that determines a weak entity is not interpreted in the same way as a regular relationship (such as the one in Model 1).

Question 4 (12 points). You are given:
Relation $R(A, B, C, D, E, F, G)$ with $F = \{AC \to EF, EG \to A, BF \to CD, CG \to D, CE \to G\}$

and decomposition: $R1(A, B, C, E), R2(A, B, E), R3(C, E, F, G), R4(A, B, C, D)$.

Determine whether this decomposition is lossy or not using the Chase Decomposition algorithm. Show your work.

Solutions:

a	b	c	d1	e	f1	g1
a	b	c2	d2	e	f2	g2
a3	b3	c	d3	e	f	g
a	b	c	d	e4	f4	g4

Apply $AC \to EF$:

a	b	c	d1	e	f1	g1
a	b	c2	d2	e	f2	g2
a3	b3	c	d3	e	f	g
a	b	c	d	e	f1	g4

Apply $CE \to G$:

a	b	c	d1	e	f1	g
a	b	c2	d2	e	f2	g2
a3	b3	c	d3	e	f	g
a	b	c	d	e	f1	g

Apply $CG \to D$:

a	b	c	d	e	f1	g
a	b	c2	d	e	f2	g2
a3	b3	c	d3	e	f	g
a	b	c	d	e	f1	g

Apply $EG \to A$:

a	b	c	d	e	f1	g
a	b	c2	d	e	f2	g2
a	b3	c	d3	e	f	g
a	b	c	d	e	f1	g

Apply $AC \to EF$:

a	b	c	d	e	f	g
a	b	c2	d	e	f2	g2
a	b3	c	d3	e	f	g
a	b	c	d	e	f1	g

Since the first row is without a subscript, this decomposition is lossless.

Question 5 (16 points). Create an Entity-Relationship diagram for the following database, capturing all the requirements below precisely. Make sure you list all the relevant attributes, underlining the keys. For each relationship, mark the participation constraints clearly (one-to-one, one-to-many, or many-to-many). If you do not find a natural key for an entity, feel free to add an id attribute.

You are creating a database that will be the next big social media hit. It is for audio only.

In this database, you store users. Each user has an email, name, password, a phone number, rank of importance and whether active or not. Emails are unique in the database. A user may invite another user. For each invitation, there is a specific date. Users may invite many users, but each user is invited by a single user. A user may block zero or more users, and users may be blocked by zero or more users.

There are topics. For each topic, there is a unique name. Topics also have a description. Topics can be related to zero or more topics.

There are rooms, identified with an id. Each room has a name and a moderator which is a user.

There are events. Events take place on a specific start and end date and time. Event have titles and rating. Each event takes place in a specific room, but rooms can have many events. On a given start and end date and time, there may be many events but only one in a specific room. Each event has one or more topics and each topic can be related to multiple events. For each event, there are users who are participants in a discussion. For each event, there are also users who are listeners. Participants can be many for each event, and events can have many participants. Listener can also be many for each event, and events can have many listeners.

There are recordings. For each recording, there is an id, duration and an audio file. Each recording is for a specific event.

Please draw your answer in the box on the next page.

Solutions: See next page. Note that the items in red were interpretable in multiple ways, so we allowed both interpretations.

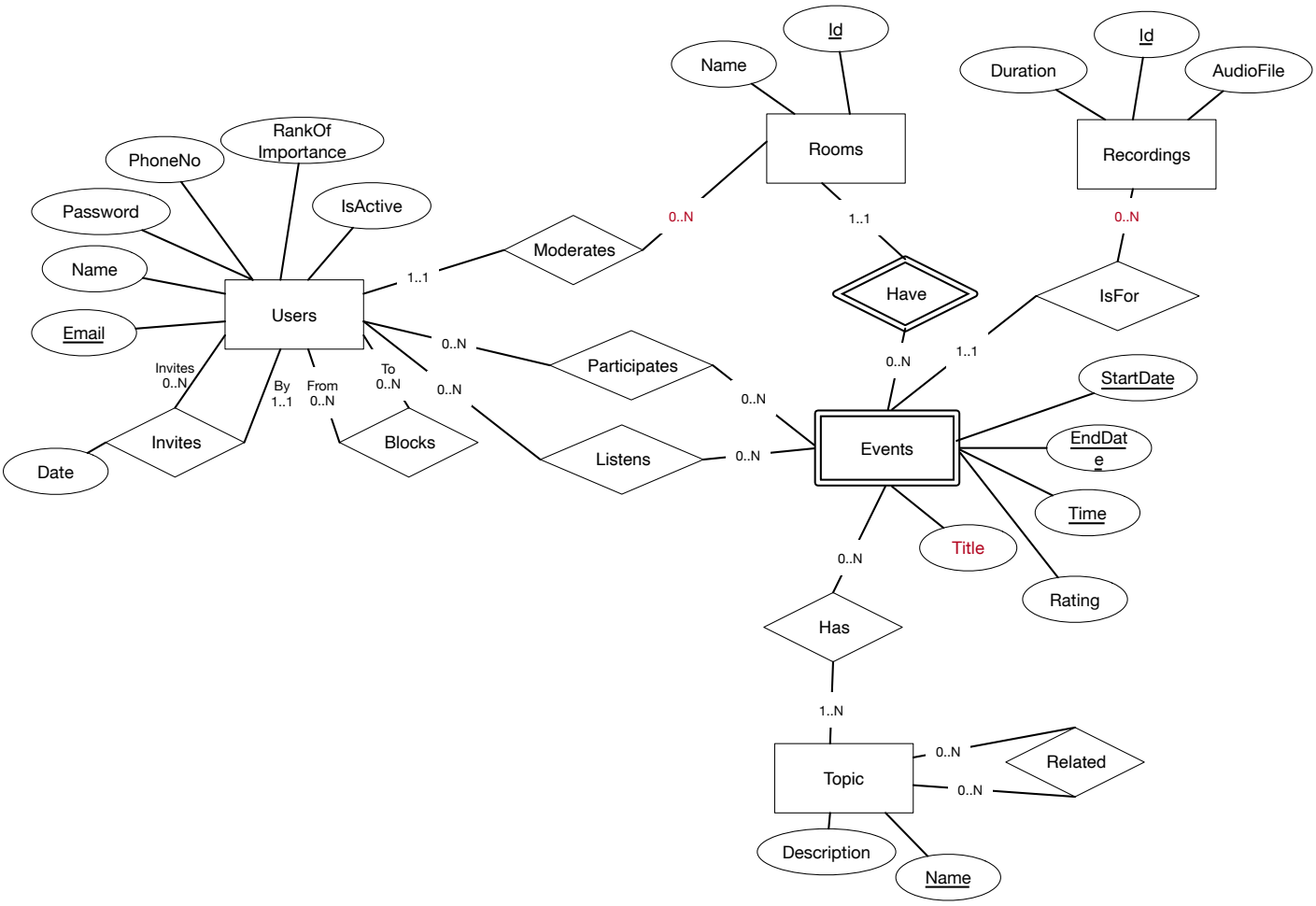


Figure 1: ER Diagram for Question 5