University of California, Los Angeles
Department of Computer Science

Computer Science 143 **Prof. Ryan Rosario**

Homework 1

Due Friday, April 9, 2021 11:59pm ***via CCLE***

For Part 1, use the following scenario for context:

Bay Area Rapid Transit (BART) is a subway system that stops at various places on the Bay Area Peninsula, City of San Francisco, underwater to Oakland and the East Bay. We have two relations in this dataset: `Station` and `RideCount`. `Station` has data about BART stations and `RideCount` contains counts of passengers that travel between two stations per hour, per date.

```
Station(Abbreviation, Location, Name)
# Abbreviation is a 3 letter code given to the station, just like an airport code.
# Location is the name of the city where the station is located. There may be multiple
#    stations in one city.
# Name is the name of the station and each station has a different name.

RideCount(Origin, Destination, Throughput, Date, Hour)
# Origin is a 3 letter code (like Abbreviation) that specifies where a trip started.
# Destination is similar to Origin and Abbreviation but specifies where a trip ended.
# Together, Origin and Destination form a "route" or "ride."
# Throughput is the number of people that traveled that "route" during a particular hour on a
# particular date.
# Date is the calendar date corresponding to the count in Throughput.
# Hour is the hour of the day (from 0 to 23) corresponding to the count in Throughput.
```

In your work, you may wish to abbreviate the names of the relations and attributes to save space. For `Station`, you can use `S(A, L, N)`, and for `RideCount` you can use `R(O, De, T, Da, H)`.

Part 1. Keys, Keys, Keys

Exercises.

(a) Find all superkeys of `Station`.

(b) Which superkey(s) of `Station` is/are candidate keys? Explain why.

(c) Which candidate key would you choose to be the primary key of `Station` and why?

(d) Which attribute(s) would you choose to the candidate key for `RideCount`? if there are multiple, which would you choose to be the *primary key* of `RideCount`? While you are free to use the algorithm described in lecture to first find all superkeys, try to instead infer the PK by the context rather than constructing all superkeys.

(e) Are there are any foreign keys in `RideCount` or `Station`? What are they?

Part 2: Schema Diagram

Suppose you are working for the data team at *Bird Scooter*, a Santa Monica based startup that aims to revolutionize how people get around on wheels. How the Bird Scooter service works: a user installs an app on their phone and enters their credit card information. The app shows a map of deployed scooters nearby. The user scans a QR code on the scooter using the app, which activates the scooter for use and begins the clock for per-minute and per-use billing. The user rides the scooter for a distance, for a certain number of minutes. When the user is done with the scooter, he/she leaves it somewhere, and marks the trip as complete in the app.

We have the following relations for managing scooters and users:

```
Scooter(scooter_id, brand, is_online)
# scooter_id is some value that identifies a scooter.
# brand is a code that represents what type of scooter it is.
# is_online just indicates whether or not the scooter is able to be used.

User(user_id, ccnum, expdate, email, name)
# user_id is some value that identifies the user
# ccnum is the user's credit card number (hashed, hopefully!) and may be null, and may change.
# expdate is the credit card expiration date and may also be null, and may change.
# email is the user's email address, and may change.
# name is the user's name

Trip(trip_id, user, scooter, start_time, end_time)
# trip_id is some value that identifies the trip.
# user represents the user that took the trip.
# scooter represents the scooter that the user rode for this trip.
# start_time and end_time represent the start and end times of the trip.
```

Exercise. Using these relations, draw the schema diagram. For an example of a schema diagram, see Figure 2.8 in the text (page 47), or the end of the Intro to the Relational Model section in the lecture slides for Lecture 2. Clearly denote primary and foreign keys (note that primary keys were intentionally withheld in the specification above).