CS 455 Principles of Database Systems

Homework 1: Relational Data Model

Name:

Instructions: Print this assignment using single-side pages. Fill in your name above, and write in the solutions in the space provided below each question. You are allowed to use the back of each page. If you used any scratch paper to show your work, append those to the end. **Note:** It is important you use this format for gradescope.

Submission: After you've filled in the answers, scan all pages into a PDF, and submit to canvas.

Problems

1. Given a relation schema $R(A_1, A_2, ..., A_n)$, what is the minimum and maximum number of superkeys that R can have? Explain.

2. What is a relation schema? How is it different than a relation instance?

3. Given a relation schema $R(A_1, A_2, ..., A_n)$, what is the maximum number of relation instances, *not tuples*, that R can have? Explain. (Recall that the domain $dom(A)$ of an attribute A is defined to be the set of allowable values of A).

4. The relational data model is not the only data model in use today. For instance, JSON widely used in emerging web applications. (1) give an overview of JSON's file structure, and compare it against the relational data model. (2) Furthermore, comment on each of the following:

   (a) Like relational databases, there keys in JSON. Are there any differences from the keys you learned about in the relational model?

   (b) Do users need to define schemas in JSON? Depending on your response, what problems may this lead to?

   (c) When might you prefer to use JSON over relational databases?

5. Consider the following relations for a database that keeps track of auto sales from a car dealership where prices can vary greatly based on the car's condition and age.

```
car(vin, model, make, price)                    options(vin, optionItem, price)
sales(salespersonID, vin, date, price)          salesperson(name, email)
```

Here are some assumptions to go on:

- Dealerships often have multiple **cars** that have the same make (e.g., Toyota) and model (e.g., Prius), so cars can only be identified by their vehicle identification number (`vin`). Each car also has a base price, which does not include any options (see below) that can be added.
- Some cars on the lot are equipped with additional **options** (such as a sunroof, spoilers, or high-end LED bulbs). Each option item carries a different price per car.
- Each **salesperson** can be identified by their email address.
- To record a **sale**, we store references to the salesperson, the car, date, and the price that it ultimately negotiated and sold for. Note that the same car may be sold more than once, even by the same salesperson. (This happens when the car is traded-in, or is sold back to the dealership at a later date.)
- To avoid scams, the dealership has a standing policy that prohibits the same car to be sold more than once on the same day.

Below each of the following relations, (1) list its *primary key* and (2) list any *foreign keys* that exist. For each foreign key $A_1, ..., A_k$ that originates in relation R and points to an attribute set $B_1, ..., B_n$ in relation S, write it as $R(A_1, ..., A_k) \rightarrow S(B_1, ..., B_k)$. Note that this implies $B_1, ..., B_k$ forms the primary key in relation S.

- car

- salesperson

- options

- sale

6. The car dealership from the previous question hired a consulting group to optimize their database. The consultants suggested an alternate database schema to use instead. They collapsed all the data down to only two tables. Critique this new schema. Compared to previous schema, list as many problems as you can spot.

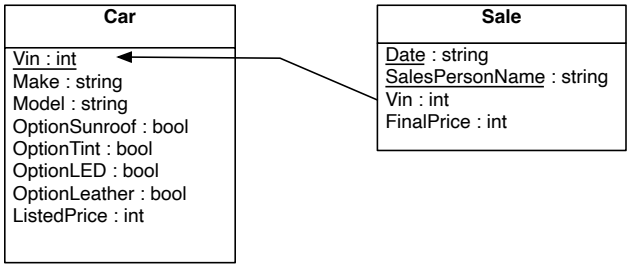

Figure 1: Newly Proposed Schema