

CS 122A: Introduction to Data Management – Spring 2019

Homework 1: E/R Modeling (100 points)

Due Date: Friday, Apr 12 (5:00 PM)

Submission

All HW assignments should be turned in with a filename that contains both your student ID and your name (e.g., 12345678_John_Doe.pdf) and must be submitted online, as a **PDF** file, through the associated (HW1 in this case) assignment in **Gradescope**. See the table below for HW 1 submission opportunities. Note that after the last deadline, Saturday April 13th, no further HW 1 submissions will be accepted at all. That is, we will not accept assignments after that time since we will be publishing the solution at that time for everyone to see. Please turn in all of your work on time! If possible, save your one dropped assignment for the end of the term when you are most likely to want/need it. (Better yet, don't use it, and you can learn everything hands-on and also maximize your best-7-of-8 score results!)

Date / Time	Grade Implications
Friday, Apr 12 (5:00 PM)	Full credit will be available
Saturday, Apr 13 (5:00 PM)	20 points will be deducted

E-R Schema Design [100 pts]

In his recent blog on “Empowering Personal Health” (<https://ngs.ics.uci.edu/empowering-personal-health/>), UCI Professor Ramesh Jain offered the following vision:

What if an app could guide patients to their health goals, similar to how GPS navigation directs people to a desired destination? What if we could use real-time information from smartphones and other sensors to redirect people around a disease, similar to rerouting drivers to avoid traffic? What if that real-time information included the latest medical knowledge as well as local and environmental conditions, empowering people to make healthier lifestyle choices?

By combining emerging technology with well-established cybernetic principles, this navigational approach to healthcare could soon be a reality. Just as mobile phones empowered personal communication and, combined with GPS technology, revolutionized navigation, a healthcare navigation system could empower personal health, completely transforming the current “disease care” industry into a patient-focused system of true healthcare.

Inspired by this vision, a group of recent UCI ICS alumni have decided to start a company to create such an app to help people take control of their health and fitness through personal health logging. Realizing that forcing oneself to improve one's health demands serious self-discipline, they have appropriately named their **Personal Health Logging** application “PHLog” (pronounced like “flog”) and named the company phlog.com. Due to skyrocketing housing prices up in Silicon Valley, the company's investors have advised basing the new startup in Aliso Viejo, near UCI, to attract employees who

might want an alternative to NorCal living expenses and traffic. As a result, being nearby, you've just been offered an internship to help PHLog.com's founders come up with a database design to underlie their new app.

Like many apps, PHLog will be centered around its *users* (the PHLoggers). There will be registered devices that are health *observers* – devices like fitbits, smart watches, sensors in mobile phones, glucose monitors, or sleep monitors – that will be a source of various kinds of *observations* (much like IoT sensor readings) about a PHLogger's health state. The app will also employ machine-learning based logic that uses the observations to identify *events* related to the PHLoggers' health activities and/or health state (e.g., periods of sleeping, walking, eating, ...). Because it helps to have support when trying to achieve a difficult goal like self-health improvement (e.g., to stick to a diet or an exercise plan), PHLog.com will enable its users to optionally participate in *interest groups* around topics (e.g., diet, exercise, depression, HIV, alcoholism, diabetes, asthma, ...). Users will be able to post *thoughts* that anyone in their interest group(s) can see. In a preliminary database design interview with the PHLog.com founders, they have shared the following thoughts on the data model needed for their new app (and its backing web site and database):

1. Each *user* will get a unique user id from PHLog (called their "phlid") when they first join the service. Users who are actually using the PHLog app for their own health logging, PHLoggers, will have a name, an email address, a residence address, and one or more contact phone numbers. Being a US-based service, at least for now, the residence address should consist of a street address, a city, a state, and a postal code. They will also have a password (hashed into a binary string value for storage). Users who are registered with PHLog just to access data (in support of PHLoggers) will just have an e-mail address and a primary phone number plus a role (e.g., "doctor", "nurse", or "trainer"). Each user has to be either a PHLogger, or a Supporter, or possibly both.
2. Each *observer* (source of data) will have a unique observer id and a kind (e.g., smartphone, smartwatch, glucose monitor, ...) Observers will also have a manufacturer (e.g., Apple) and a model (e.g., "Apple Watch Series 4") and an optional software system name and version (e.g., "IOS 12.2").
3. Each *observation* will have a unique observation id, an observation start time, and an (optional) end time (in case the observation is from a time interval rather than a specific point in time). Different kinds of observations exist in the PHLog universe and have additional information. A *heart rate* observation also has a rate (an integer value). In contrast, a *blood pressure* observation adds two integer values, called the diastolic and systolic pressures. A *picture* observation instead adds an image (a binary value) with an accompanying label (text) and description (also text). Finally, a *PHLogger* observation adds just an ad hoc comment (text) that the PHLog user might wish to include in their health log (e.g., "Really feel like crap right now.") for later reference.
4. Each *event* has information including a unique event id, an event name (e.g., "sleeping", "walking", "eating"), and a start time and (optional) end time. It may also optionally have an origin value (text) that the PHLog app can use to record whatever it wants about the nature of the logic (if any) that led to the recording of the event (e.g., the name of the algorithm that produced it).
5. Each *interest group* has a unique name, a topic (e.g., "exercise" or "cancer"), a textual description, and the date when it was created.
6. Each *thought* will have a thought number (more about that later), a text value (which is the thought itself, e.g., "Week two of chemotherapy definitely sucks!"), and the time when it was entered into the service.
7. Each observer is *owned* by a PHLogger. An observer has one owner, but a PHLogger can own several observers.
8. A given observation is either *observed* by an observer or *recorded* (manually) by a user. (Note that a PHLogger observation will not be associated with an observer.)

9. Events are things that have *happened* to a PHLogger. A given event is associated with a specific PHLogger, who may (of course) have many associated events.
10. Optionally, an event may be *indicated* by one or more observations (e.g., the set of observations that led the application's algorithm(s) to decide that this event occurred, like the observations leading to a conclusion that a PHLogger was asleep for a particular time interval). Note that it is up to the app to decide when it does/doesn't want to explicitly record a given event's indicators (hence the "may").
11. Thoughts are *offered* by PHLoggers, and their thought numbers capture the order in which they were thought of by their offering thinker.
12. A thought must be *about* the topic of one or more associated interest groups (e.g., a thought about "cancer", or about "cancer" and "diet").
13. Last but not least, a PHLogger can be a member of zero or more interest groups.

Design an E-R schema to represent the required information and express your design in the form of an E-R diagram. (Please use the E-R notation from either the lectures or the textbook; do NOT invent your own notation or use UML. You can find a decent E-R model tutorial on the web at http://www.tutorialspoint.com/dbms/er_model_basic_concepts.htm, but beware of its slightly different ISA notation -- its ISA notation closest to ours is specialization -- which you should **not** use here because its ISA triangles are upside-down/ambiguous and it doesn't address covering/overlapping constraints.) Be sure that your E-R design captures all of the implications of the PHLog user model and business model, including the aspects below ((a)-(b)), and be sure to use the advanced E-R modeling goodies such as weak entities, ISA relationships, composite attributes, set-valued attributes, etc., where appropriate. Please use the book's/slides' ISA notation and be sure to think about overlap and covering constraints when detailing your design. NOTE: Do NOT attempt to design your E-R schema by making just one pass over the bullets above! You will probably need to carefully analyze all of the requirements together in order to identify the available attribute and relationship commonalities for ISA purposes; attempting a one-pass design is likely to lead you to miss important (inheritance-based) sharing opportunities and lead you to a messier-than-necessary design. Cleanliness matters!

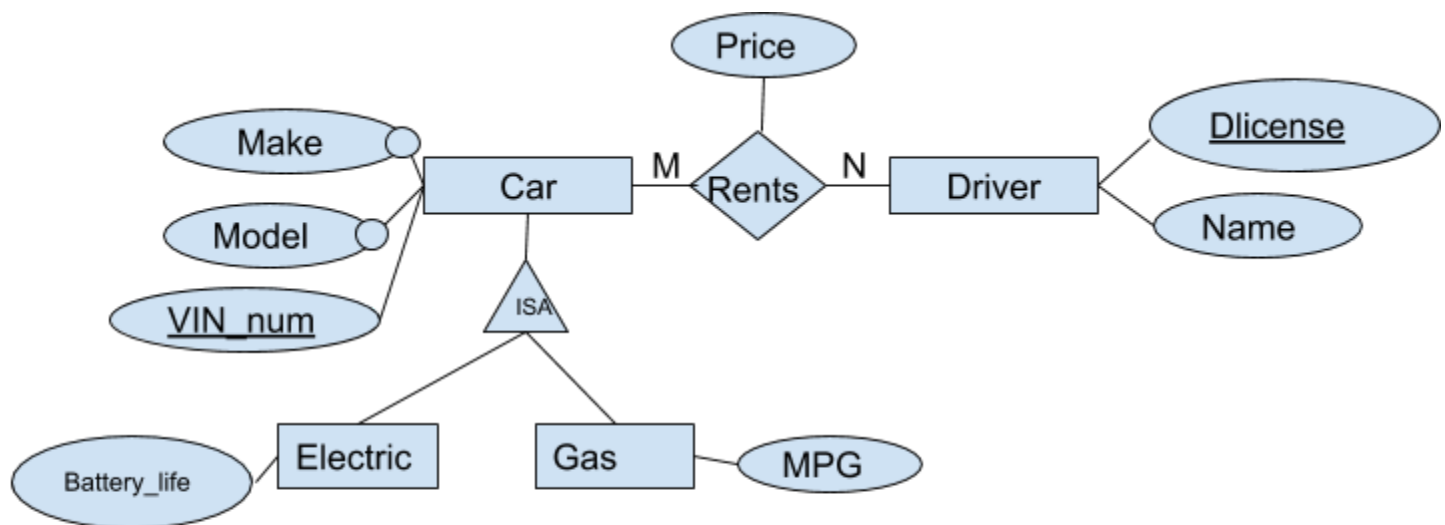
Using the schema template that we will provide, draw your E-R diagram, including:

- (a) [50 pts] All of the relevant entities and their associated attributes, including their keys. Nearly all of the entities needed will be in the template (with one possible exception). You should **NOT** introduce more than one additional entity or use fewer entities than those given.) Also, none of the entities in the template are marked as weak; if you wish to change that, you may.
- (b) [50 pts] All of the relevant relationships and their associated relationship attributes, if any. For the relationships, each one must also have appropriate cardinality and participation constraints.

Be sure to download the HW #1 PDF schema template from the Attachments area of the CS122a web page and use that as the basis for drawing the E-R schema that you turn in. (Your solution will **NOT** be accepted if you do not use our provided layout for your diagram! Without such standardization it becomes too difficult to grade everyone's answers with sufficiently thoroughness.) Please use some form of software to draw your E-R diagram - but please, do **not** change the spatial placement of the entities - instead of drawing it by hand. Again, you must use the concepts and their associated notation from class (see the lecture slides!) - there are many variations of the E-R model out there, but we

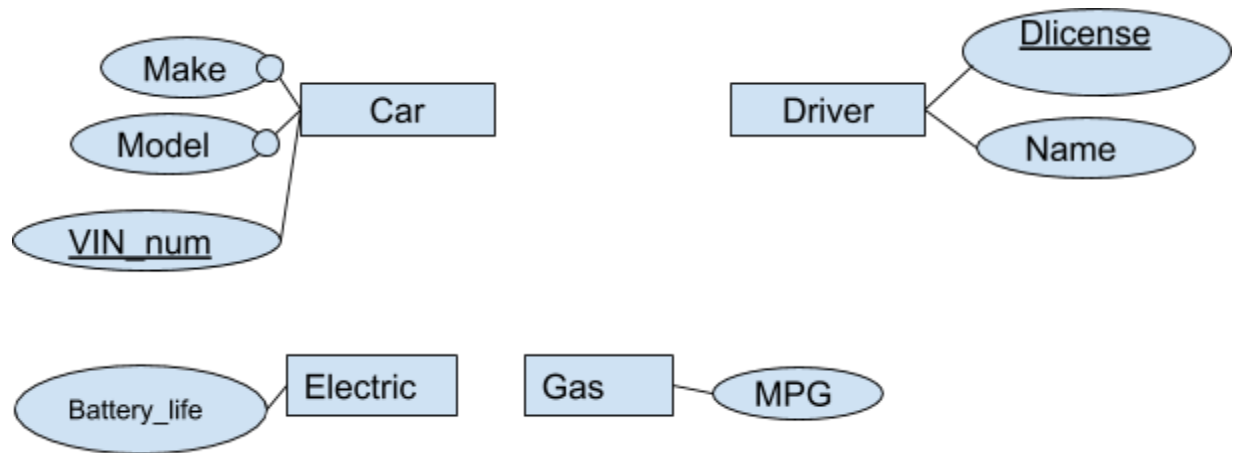
are going to have everyone use the same (ours!) variant for this quarter. You can pick and choose your own favorite later, if you end up building E-R models in your future job life.

Important: As you will see in the templates, there are two similar pages that have all necessary entities inside. **For the first one, it should only contain entities and their attributes without any relationships. For the second part of your overall E-R diagram, just show the relationships, their cardinalities, their attributes (only the relationship attributes), and the ISA partition constraints.** Please do not include any entity attributes in the second E-R diagram (You will lose points!). As an example, suppose that your final design for PHLog.com's conceptual schema is the following E-R diagram:

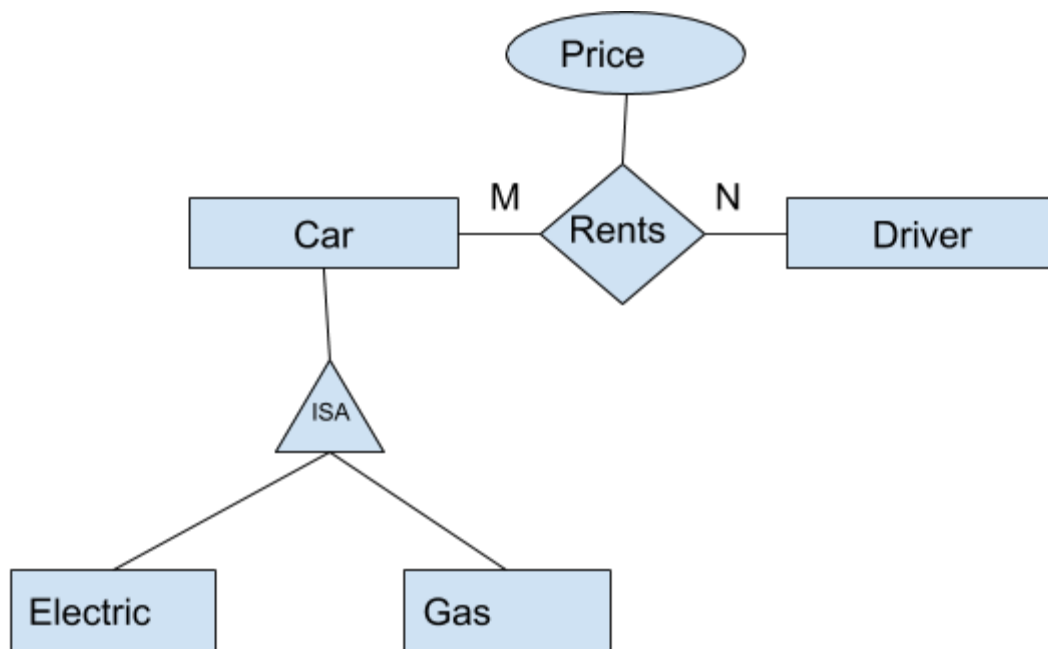


Electric cars OVERLAPS Gas cars (e.g., to model Hybrid cars)
Electric AND Gas cars COVER Car (i.e., there are no "just cars")

In your first E-R diagram you should show all of the entities with their attributes but no relationships. This diagram will be used to check that you have correctly included all of the attributes needed for PHLog.com.



Then, in your second E-R diagram, you should show the graph of all entities (without attributes) and relationships (with attributes if any):



Electric cars OVERLAPS Gas cars (e.g., to model Hybrid cars)
 Electric AND Gas cars COVER Car (i.e., there are no "just cars")

This diagram will be used to check your overall design w.r.t. its relationships -- and it will give you a nice overview to refer back to later on in the project.

Note that if the above pictures are indeed the same as your own final conceptual schema design for PHLog.com, you

should probably read the founders' information again more carefully (😊).