Part 1: Software Setup

- Create or turn on a Windows VM on your vsphere account.
- Go to mariadb.org and download MariaDB Server. Follow all the defaults for the install. When you're prompted to create a password for the root user, choose something that you'll remember.
  - It should come with the HeidiSQL client as well. It's a bit ugly, but it will work!
- Download and Install Git for Windows, recall that files related to labs (not today) will need to be cloned or pulled from Git. We'll also need to create an store an SSH key for our account, you can follow these directions to do so.
- Finally we're going install Anaconda, which is a python distribution used for data science and analytics.
  - As we move forward in labs, I may have .py files to execute and populate the database with data.

Part 2: Data Modeling

Overview

You just started a new job a Habitat for Huge Manitees. The mission of this international organization is to raise money to preserve and raise awareness about the ecosystems in which Manitees live. They want to preserve healthy (and huge) manitees as the world grapples with climate change.

The organization has been around for some time and has a large base supporters that give donations, both one-time and sustaining, subscribe to organizational email newsletters and podcasts and occassionally sign petitions in support of specific causes related to manitees. The organization also recieves grants from other organizations to provide financial support for their projects targetting specifc geographic populations of manitees.

They're embarking on migration of their technology platforms because the existing platforms became to expensive while also not meeting the needs of the users. The data that is captured about the supporters and their activities is critical for generating insights that inform the strategic direction of the organization.

While data modeling shouldn't be done in a bubble, this lab is meant to start to get you familiar with how this process might go.

Business Information

Unified Action Center - Habitat for Huge Manitees has a seperate and secure system for storing user login information, but this login information ties into a supporter record based on a unique login_id (a UUID). We don't need to

worry about this system except for this login_id. Once a user is logged in to our website, they can take any number of actions. A login_id can be linked to one or more supporters because a shared login implies a relationship.

Supporter - a person or organization that contrbutes to Habitat for Huge Manitees through donations, subscriptions or petitions. Email addresses are typically used to differentiate these supporters. We need to know when our supporters first joined us, which country they are from and their first name. Knowing their last name and birthday is helpful, but not necessary. Importantly we don't want to capture data for minors (those under 18).

Donation - this refers to a monetary gift to Habitat for Huge Manitees. Because of the organization's international precense, gifts are typically given in a local currency and require that their currency to be converted. Donations tied to a specific supporter and must at minimum track the date, the US Dollar amount and the original amount. We also want to keep track of payment methods.

Sustaining Donation - these donations are given once per month on the day corresponding to when the donation was intitated. Sustaining donations are not always active, that is they may have been cancelled, but for each month they were active they will have one and only one related donation record. For sustaining donations we need

Email Newsletters - supporters can subscribe to newsletters that cater to their chosen interests, of which they can have many. Habitat for Huge Manitees supports newsletters in several languages. Supporters are not required to subscribe to newsletters. It's critical for us to understand the URLs from where these supporters subscribe. Most obviously, a supporter must have provided us with an email address.

Podcasts - Habitat for Huge Manitees offers several different podcasts, some remain active, while others remain available without new episodes. We offer our podcasts in several languages and allow our supporters to subscribe to them from our website or a podcast app on their phone. It's important for us to know which podcasts they subscribed to, when they subscribed/unsubscribed from a podcast, the date of these events and whether it was done through our website or an app.

Petitions - on our regular basis we publish and reach out to our supporters and ask them to sign petitions in support of issues related to Manitees and their habitats. Sometimes these petitions are for everyone, while other times they are associated particularly countries that are proximate to the habitats. These petitions are entirely optional, but a user typically provides their email address, first name and last name during the signature process.

Data Modeling in Action

1) Create a list the entities that should be included in our data model. Remember that entities often refer to people, places and things.

2) Now that you've come up with the entities, let's start to create an Entity Relationship Diagram (ERD). For this exercise, we'll focus on the entities and relationships, rather than diving into the attributes. You may use Crow's Foot or Chen's notation. Draw out your entities as boxes, connect related entities with a line, depending on your notation use a diamond (or just a label) to note the verb that describes the relationship(s). Appropriately denote the type of relationship you see between the entities (1:1, 1:M and M:N).

3) With the information that you've been given, are you able to identify the cardinality of any of these relationships? If so, please do so.

4) List a few questions that you have related to the entities or the relationships that you need to know in order to ensure this data model will meet the needs of the end users. In other words, is there any ambiguity that is already posing challenges? Take some time and be thoughtful, our role as technologists requires us to ask questions so that we can build the right things!

5) We won't include them in our ER Diagrams yet, but let's start to map out the attributes of our entities. You can draw rectangles with the entity names at the top and attribute names underneath. Leave extra space on the right for writing out data types next lab. Also, put an asterisk followed by the first letter of the constraint type(s) near any attributes that you think may require a constraint of some kind (unique, check, notnull or default). So a unique constraint would be denoted as*u.

6) Now that we know entities, attributes, relationships which attributes are candidate keys? Which would you select for the primary keys, how about foreign keys?

When you're done, please turn in your lab and then you are free to go.