

CS 327E Lab 1: Relational Database Design

Learning Objectives:

1. Work with real-life data
2. Work in small teams
3. Create data models that describe your data
4. Design and create a relational schema in MySQL
5. Get some exposure to LucidChart and Github

Prerequisites:

1. Lab partner
2. Personal github account
3. Private repo for your team
4. Git client installed on your laptop
5. MySQL server installed on your laptop
6. Optionally: MySQL Workbench installed on your laptop
7. LucidChart account with education upgrade

Steps Outlined:

1. Choose 1-3 interesting and related datasets. The datasets must be both structured and in one of the following formats: csv, json, xml. Chapter 6 from our Data Wrangling text discusses various datasets that are available from <https://github.com/jackiekazil/data-wrangling>. You are not obligated to use the datasets mentioned in the book, but you must use a dataset that is not proprietary. Once you have chosen your datasets, create a text file and list each dataset with a brief description for each one. The description should include the name of the organization that published the dataset, the file format of the dataset, as well as some of the attributes that you found of interest and plan to model in the database. Be sure to include the link to the web site from where you obtained the dataset. Also, if you have noticed any anomalies in the raw data, be sure to mention those as well as any challenges you foresee when you start working with the data. Name the file data_sources.txt.
2. Draw a single Conceptual Model for your dataset(s). If you're working with multiple datasets, you must create a single unified Conceptual Model for the data. For example, if one dataset contains flight delays and another dataset contains weather events, you must design a model that joins these two datasets. The model should have a minimum of 4 entity classes and 3 relationships. Use LucidChart to create your conceptual model. Download the file as a pdf and name it conceptual_model.pdf.
3. Provide a brief description for each Entity Class and its attributes. These descriptions should be in an excel spreadsheet with the following column headings: **Source** (column A), **Name** (column B), **Type** (column C), and **Description** (column D). The Source column should identify the data source in question. The Name column should contain the name of the Entity Class or attribute. The Type column should contain either "Entity Class" or "Attribute", depending on what is being described and the Description column should contain a brief description of the Entity Class or attribute. The attributes should be

organized by Entity Class, such that if attribute A belongs to Entity Class C, the description for A should follow the description for C. Save the file as data_dictionary.xlsx.

4. Derive a Logical Model from your Conceptual Model. Translate each Entity Class into one or more relations or tables. Translate each many-to-many relationship into a junction table. The resulting model should be in at least 1NF. The Logical Model should specify the datatype for each attribute as well as the primary key for each relation. The model should also specify all the required foreign keys. Use LucidChart to create this model. Download the file as a pdf and name it logical_model.pdf.

5. Write the CREATE TABLE statements for each relation that is in your Logical Model. Be sure to define all the proper constraints including primary key, foreign key, not null, and check constraints. Put all the create table statements in a script called create_tables.sql. Make sure that the create table statements are in the right sequence such that the parent tables precede the child tables. The script should also create a database for storing the tables as well as drop the database if one already exists. To drop and create the database, use the commands:

```
drop database if exists <database name>;
create database <database name>;
create table <table name>
...
```

Run the create_tables.sql script on your local MySQL instance either through Workbench or from a MySQL shell. Fix any syntax errors you encounter and re-run it until it functions as expected.

6. Save all files in a folder called lab1. Add this folder and its contents to your team's github private repo. Locate the **commit id** that you will be using for your submission. This is a long 40-character that shows up on your main Github repo page next to the heading "Latest commit" (e.g. commit 4c91ea6dd1a04b12e77c244d49454d77708dd61a). Copy the commit id and paste it into an email. Your email should contain the following: a link to your team's repo and the commit id of the code that you want to submit. Your email should be addressed to the professor and both TAs and carbon copying your lab partner. The email should have the following subject, replacing <TeamName> with your actual team name: [CS327E][Lab1][<TeamName>]. The email is **due Tuesday, 09/27 at 11:59pm**. If it's late, there will be a **10% grade reduction per late day**. This late policy is also documented in the syllabus. Note: Only one person per team should send the submission email.

Naming Conventions:

1. The first letter of an Entity Class should be capitalized. If an Entity Class contains more than one word, use an underscore between the words and capitalize the first letter of each word. For example: Class_Enrollment.
2. All attributes should be in lower case. For example: eid.

Diagram Conventions:

1. Use the single-column Entity shape for the conceptual diagram, which only has a textfield named Field.
2. Use an arrow to designate a subtype-supertype relationship. The arrow should point to the supertype.
3. Use the three-column Entity shape for the logical diagram which has 3 textfields named Key, Field, and Type.

Teamwork:

1. We will use 3 class meetings (09/19, 09/21, and 09/26) to work on this lab.
2. We expect you and your partner to divide up the work as evenly as possible and to both equally contribute to the assignment.
3. We want you to use the Github Issue Tracker to assign and track the status of tasks.

Resources:

Lab 1 Setup Document: <http://tinyurl.com/zwmlg6o>
Lab 1 Grading Rubric: <http://tinyurl.com/j54j87d>
Lab 1 Team Sign-up Sheet: <http://tinyurl.com/j6hzgww>
Sample Conceptual and Logical Models: <https://github.com/cs327e-fall2016/snippets>