

**CMPSC 312
Database Systems
Spring 2022**

**Lab 4 Assignment:
Schema Design**

Submit deliverables through your assignment GitHub repository.

Objectives

To design databases after evaluating the different levels of schema. To determine the natural relationships in the data to be used to build connections between tables. To evaluate some of the ethical considerations involved with database design.

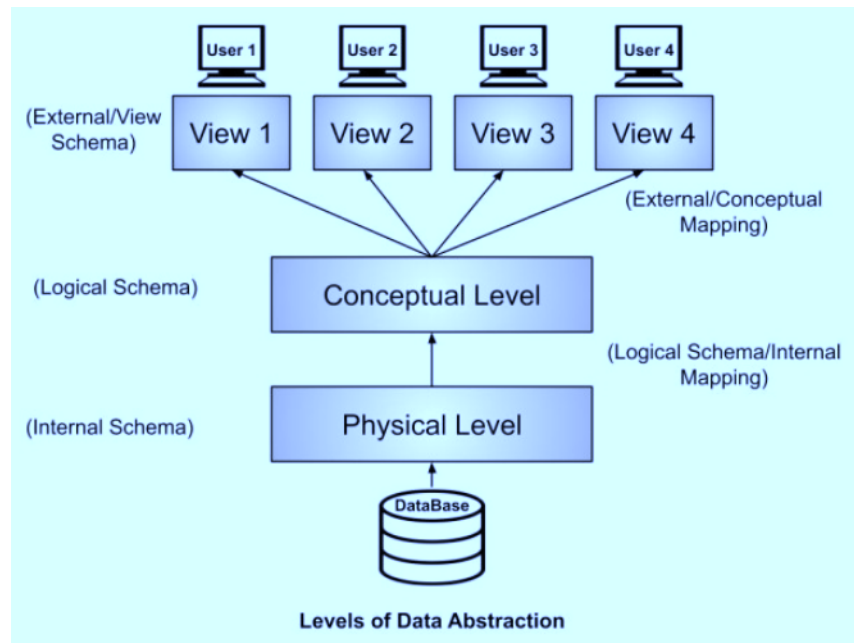


Figure 1: There are different levels of schema involved in designing a database. Each level helps the architect of the database to consider degrees of functionality, maintenance, and user-interactivity of the system.

GitHub Starter Link

<https://classroom.github.com/a/ciEJ3mgi>

To use this link, please follow the steps below.

- Click on the link and accept the assignment

- Once the importing task has completed, click on the created assignment link which will take you to your newly created GitHub repository for this lab,
- Clone this repository (bearing your name) and work locally
- As you are working on your lab, you are to commit and push regularly. The commands are the following.

```
- git add -A  
- git commit -m 'Your notes about commit here'  
- git push
```

Introduction

The design of a database is very important to its intended functionality. In addition to the relationships that must exist between each table, there are other considerations, as well, which must be made to ensure better functionality. In addition, during the design of a database, an architect may include thinking to facilitate the maintenance of the database. Discussed below are the different levels of a schema which encourage thinking about functionality and maintenance.

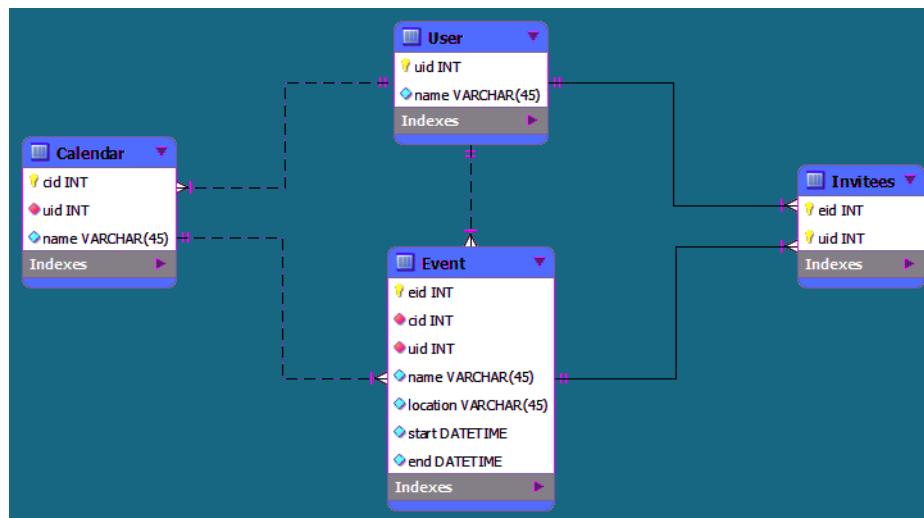


Figure 2: Logical Schema: A bird's eye view of the database structure. This design allows one to determine the relationships inside and outside of the tables. This schema is represented using an Entity-Relationship Diagram (ERD). Here, rectangles represent entities (e.g. tables), ovals, attributes (e.g. columns), and diamonds, relationships (e.g. one-to-one, one-to-many, many-to-one, and many-to-many).

Schema

When designing a schema, there are different levels in its design. In Figure 1, we note these levels in a schematic drawing of a typical database.

- **A View Schema or External Schema**

- A View Schema defines the design of the database by what it will *look like* when deployed. At this level, the designer can decide how the user is to understand this database as a tool. This design also dictates how the database will be presented to the user – by a command-line driven application, by a browser and etc. This level also determines the level of necessary training or skill in terms of programming that would be necessary to interact with the system.

- **The Logical Schema or Conceptual Schema**

- The Logical Schema, shown in Figures 1 and 2 defines the design of the database at the conceptual level of the data abstraction. Here, we note the the definitions for tables, attributes, entities, constraints, relationships and similar which serve to connect each table to the others. In addition, we also note here that each entity in any of the tables has been carefully placed according to a logical relationship.

- **The Physical Schema**

- This is the *hardware* level of the schema that describes how the data will be managed using computer equipment at the physical level of data abstraction. At this level, one notes how the data will be *physically* stored on the storage device(s) such as cloud technologies, local-hard drives, and similar. The kind of computer system (in terms of memory, CPU power and similar) would also be clear from an evaluation at this level.

What To Do

Locate and obtain a source of data from an online source of data. A list of sites that may offer data for this project is given below. **Please note; keep your database to around up to 1 Megabyte in size.** If there is additional data from the site, then you are asked to cut the data down to size in a way of your choosing (taking a subset of the data according to some rational or by cutting the data off at a particular size).

Once you have chosen your data, you must decide on the purpose of a database to use with this obtained data. In your `writing/report.md` Markdown file, you will be asked to reflect over the different levels of the schema design as you develop your system. Build your SQL database using at least three tables and return to the report Markdown file to explain your choices.

Data Sets

To build your schema, please select a data set which is available for on line for free. The following links may help to get you started in finding the data that you need to build your schema. Your database is to be composed of three or more tables.

Please use a new dataset which you have not used for any other project.

- Pelletier Library at Allegheny College (online services): <https://allegheny.libguides.com/az.php>
- World Health Organization: <http://www.who.int/>
- The World Bank: <https://www.worldbank.org/> and <https://www.who.int/ncds/surveillance/en/>
- Demographic and Health Surveys: <https://dhsprogram.com/>
- Harvest Choice: <https://harvestchoice.org/>
- Food and Agricultural Organization: <http://www.fao.org/home/en/>
- World Population Prospects: <https://population.un.org/wpp/>
- Centres for Disease Control and Prevention (CDC): <https://www.cdc.gov/>
- US Food and Drug Administration Home Page: <https://www.fda.gov/>
- The US Census: <https://www.census.gov>
- Institute for Health Metrics and Evaluation: www.healthdata.org/
- IBM's collection of opensource data sets: <https://developer.ibm.com/exchanges/data/>
- Google's opensource data sets: <https://research.google/tools/datasets/>
- Data.world: data for business-based questions: <https://data.world/>
- Kaggle: <https://www.kaggle.com/>
 - Kaggle's Star Trek Scripts (Could be a cool idea!): <https://www.kaggle.com/gjbroughton/start-trek-scripts>
- And many more that you may conveniently find using online searches. Please remember to cite your data in your report; giving the name and its web address.

Summary of the Required Deliverables

Please submit your work by pushing it to your GitHub Classroom repository.

1. **Report document:** You will modify the file `writing/report.md` to respond questions in the document.
2. **Database-building file:** You will submit your edited build file (`src/myDB_builder.txt`) to be used to build your database from your obtained data files.
3. **Data files::** Please submit all your data files in `src/data/` and your compiled database as `src/myDB.sqlite3`.

In adherence to the Honor Code, students should complete this assignment on an individual basis. While it is appropriate for students in this class to have high-level conversations about the assignment, it is necessary to distinguish carefully between the student who discusses the principles underlying a problem with others and the student who produces assignments that are identical to, or merely variations on, someone else's work. Deliverables that are nearly identical to the work of others will be taken as evidence of violating Allegheny College's Honor Code.