

FINAL PROJECT PHASE I - Databases

Phase I due date: Thursday November 16, 2023, 3:00 PM (Baltimore Time)

The final project in this class will give you the opportunity to design and implement a moderately large database system in a domain of interest to you.

Rather than specifying one fixed domain/task specification that the entire class must address, groups will be able to select a domain/task particularly relevant to their outside interests, such as an investment portfolio database, a medical database, an astronomy database, a pharmaceutical database or a small business database.

The purpose of the first phase is to encourage/require groups to begin substantial work on their projects early enough that they can finish their complete project in a timely way, as well as to identify projects in advance that are either unsuitable or not sufficiently challenging. If you have questions about either the suitability of your proposed domain or your specialized plans, please make an effort to discuss such questions in advance of this Phase I submission. The bulk of your effort here, however, is to begin the database design and implementation and the real/interesting target-domain-data identification/collection process.

In general, your final project will constitute an end-to-end system, with an interface for data input (either from the user via GUI forms, the MySQL web interface or Perl-based data-extraction from text or on-line data sources), and an interface for data output (both through the generation of reports, complex SQL queries and definitions of views to provide customized perspectives on the database for different users).

Phase I Requirements:

- (1) Who are your team members (normally a maximum of 2 people, but with prior approval teams of 3 may be allowed if their project is sufficiently complex and the grading will be based on this expectation of greater complexity and substance).
- (2) Briefly describe your target domain (e.g. a world geographic database)
- (3) Give a reasonably comprehensive and representative list of the kinds of English questions you would like your system to be able to answer (*minimum* 15).

For example, "Compute the mean literacy rate for countries with a per capita income of less than \$400/year, grouped by continent."

Please note that these queries are not the only thing you will need to support, just some basic objectives to help focus your design choices.

- (4) Design and show a relational data model that you plan to use for your system, with a preliminary implementation in standard SQL data-definition-language syntax. This specification should include appropriate primary key, foreign key and domain specifications for each relation/attribute, as well as the NOT NULL constraint when appropriate. You may also find it useful, but not required, to create a few insert-into statements that populate your schema designs with representative values (both to document your choices and to exercise them. You are welcome to change and augment your design and its specification by Phase II, but any time investment now will reduce effort later.

- (5) Submit a set of SQL statements that will implement a representative sample of your target queries, including some of the more interesting or challenging cases. This is primarily to get you to think about your design and how it will be exercised as well as any limitations, so focus on queries that would be useful for doing so, rather than creating trivial or non-insightful queries just to fill space.
- (6) Provide a plan for how you will load the database with values.
 - If you plan to extract/import data from on-line sources, briefly describe what are the sources (e.g. personal data, or provide URL's) and what are any format conversion issues you expect to encounter.
 - If you plan to input your data primarily through a web or form-based interface, briefly describe this interface and the issues involved.
- (7) Very briefly describe the form/type of output or result you plan to generate or any special user interface issues (e.g. views) that you plan to implement.
- (8) What are the specialized/advanced topics you plan to focus on in your database design? Examples include:
 - security (e.g. banking)
 - object-oriented or distributed database design/implementation issues
 - advanced SQL topics (triggers, cursors, JDBC, etc.)
 - optimization/tuning
 - data mining
 - complex data extraction issues from online data sources
 - natural language interfaces
 - particularly advanced GUI form interface and/or report generation

Students in 601.415/615 should do major work in at least one of these areas and minor work in at least one other. Students in 601.315 should do at least minor work in one of these additional areas of specialization. More details will be discussed in class.

The final project will include the full implementation of each of the points proposed or described in Phase I, including (but of course not limited to) a full suite of SQL queries to answer interesting questions in the domain. A demonstration of the system will be required, either on-line or through snapshots of sample output. Your Phase I submission will be used as part of your final project submission and thus should be included as part of your Phase II submission, including any revisions that you wish to make to update Phase I if it has changed.