

FINAL PROJECT PHASE II

601.315/415/615 - Databases

Phase II recommended submission date: Friday, December 8, 2023, 11:59 PM

No late penalty if submitted before: Wednesday, December 20, 2023, 4:00 PM

What to hand in:

The goal of Phase II of your final project is to submit the code, data and sample output/screenshots of what you accomplished, and submit a relatively straightforward writeup summarizing the primary accomplishments of what you submitted (in a separate single PDF file).

- (1) Your Phase I submission may be considered an integral part of the final project, and it will not be necessary to repeat this information in your Phase II report. Ideally you should copy the Phase I contents at an appropriate place in your final report, to make it easier for to review the information, or at a minimum attach it as a separate document at the end.
- (2) On your **cover page**, include (a) the names of your team members, (b) the section each is enrolled in (315/415/615), (c) your email addresses, (d) a 2-3 line description of your application domain, and (e) a URL where an on-line project interface is located *if* you have one (which is not necessary) or a description of how/where to view your results.
- (3) If your project description or database design has changed since your Phase I submission, describe these changes or additions.
- (4) Briefly describe how you loaded the database with values, including the sources of your data (e.g. URL's) and pointers to any specialized code you developed for pre-processing, extracting or loading the data.
- (5) If you did not use mysql on `dbase.cs.jhu.edu`, describe the exact software/hardware platform you used.
- (6) Provide a brief user's guide, describing in sufficient detail how a user could run your code, especially if using a non-standard platform or interface. If the comments in your code are sufficiently detailed, a pointer to them will be adequate.
- (7) List and very briefly describe your major/minor areas of specialization, such as (1) complex extraction of real data from online sources, (2) an interesting interface demonstrating significant accomplishment, expanded education and/or originality, (3) original and challenging data modeling (e.g. for genomic, geographic or multimedia data), (4) security, (5) transaction control, (6) natural language interfaces, or (7) data-mining and knowledge discovery.
- (8) Give an itemized list of your project's particular strengths or selling points, especially things that may not be obvious upon inspection of your code or output.
- (9) Give a brief itemization of your project's limitations and list suggested possibilities for improvement or worthwhile extension with additional time.

- (10) If your code submission includes components that were (a) written by people not on your team or (b) written by yourself for another course, for prior research and/or employment, please very clearly explain/document which components were done elsewhere. Such outside borrowing is permitted if clearly documented, and you are permitted to build upon prior accomplishment, just like you may utilize software libraries from elsewhere, but your project will be evaluated on the original work done for this course.
- (11) Include a set of reasonably formatted “output” from your project, appropriate for the focus of your project, and demonstrating interesting and challenging extraction, synthesis and summarization of information in the database and supporting your specialized topics, with sufficient titles and or notes to make it clear what the output represents. The nature of the output will depend on what is the the focus of your project.

Ample screenshots that show off all aspects of your project are strongly recommended to make sure all aspects of your project’s accomplishments are clear, and you may provide a live demo (via recorded video or in person/zoom) if you feel appropriate.

- (12) Submit a full relational table specification of your database in the SQL Database Definition Language (DDL). This specification should include both the data type of each attribute, the NOT NULL constraint when appropriate and sample data values for each attribute represented as comments. You should also specify the primary keys (e.g. PRIMARY KEY (*ssn*)) and referential constraints (e.g. FOREIGN KEY (*mgrssn*) REFERENCES *employee(ssn)*). Many groups included this specification in their project phase I writeup, but please update it if it has changed.
- (13) Include a copy of all SQL code used in your system (commented appropriately), as well as any additional code you used for data acquisition and input.
- (14) You are required to submit a complete archive of all your code (SQL/stored-procedures/python/PHP/etc.) and data files, ideally in a single compressed nested directory. Your directory should be named as follows to make it easier for me to identify when I uploaded it:
yourgivenname_yoursurname if done individually (e.g. john.doe.zip)
or
partner1givenname_partner1surname_partner2givenname_partner2surname
(e.g. john.doe_adam.smith.zip)

Please also submit your writeup as a *separate* PDF file including your names as above (e.g. john.doe_adam.smith.pdf).

In most cases projects should be submitted via the normal homework submission procedures on gradescope, but if the combined project directory size is very large it may be necessary to put your zipped submission directory onto an accessible location on your jhu public.html directory (strongly preferred) or other hosting service (please provide the URL and instructions for access). But please do **NOT** use dropbox or any repository that’s password protected or requires login.

Project teams should only submit the project **once** via either partner, and not repeated for each partner. Your coverpage and multiple names in your directory/writeup filenames will make it clear who are the project authors.