SOEN 363: Data Systems for Software Engineers

Database Project Phase N (0)

Project Proposal and Non-Functional Requirements

Fall 2024

Date posted: Tuesday, October 1st, 2024.

Date due: Tuesday, October 8th, 2024, by 23:59.

Weight: 0% of the overall grade.

Group Project. You must work strictly within your group.

Overview

This document outlines the Non-Functional Requirements of the database project, in which you implement a relational database; and subsequently in phase II, you will migrate the data into a NoSQL [1] database. Using a topic of your choice, you want to implement a relational database whose data is populated from at least two different public sources. In the project, you will design, implement, and populate your database by consuming public APIs [2]. You may write additional code for data transformation that eventually create a 'normalized' database.

NOTE: A presentation session at the end of the semester will be arrange so that you demonstrate your project for peer review.

Non-Functional Requirements

The following non-functional requirements must be meet.

- You must collect the data from at least two different public data sources. Note that you will need to demonstrate the data collection process during your project presentation.
- You must show a use of public API in your data collection. Note that many public free APIs have limited bandwidth. You must plan your data collection accordingly.
- Your database must contain a large data set. While this may be subjective, a database of size 300MB and above may be a good example of a large data set. Ideally, a couple of Gigabytes would be ideal, however most public platforms provided you with limited bandwidth. Please plan your data collection, accordingly.

NOTES:

- Use a topic other than movies (IMDB or any other similar data sources).
- While you may use any RDBM database, note that some databases may have limited or no support for user-defined types (i.e. MySQL) or perhpas triggers. As a result you may what to explore other databases e.g. PostgreSQL, MS-SQL Server, etc.
- While using cloud services are encourages, note that at the end you must submit your database in the system. The submission includes the DDL and the DML queries + the script code in zip format. Details will be posted in the project specification document. Therefore, it will be strongly recommended that you install the database server locally on your machine.

Implementation Platform

You may use PostgreSQL [3] (or any other free RDBMS implementation platforms). You may use any programming language of your choice for API consumption, data population, and database instance creation.

What to submit

The detailed requirements for the project will be posted later. However, you must finalize the topic of the database. You are required to submit a one-page proposal document that addresses the following:

- Your topic: A possible name + the scope of your database.
- Name and ID of the team lead.
- Name and IDs of the team members.
- A brief description of what will be included in the database. A list of high-level entities would be a plus.
- Datasources (at least two): The source, the API type.
- Implementation platform: proposed RDBMS and NoSQL System.
- Programming platform: Language platform (for API consumption) and/or scripting language or any other intermediate tools that you plan to use.
- Additionally include an estimated plan for the complete data collection.

Contact your tutorial Instructor before submitting your information. For any questions on forming teams, minimum / maximum team size, etc. please contact your tutorial instructor.

References

1. `https://en.wikipedia.org/wiki/NoSQL`
2. `https://github.com/public-apis/public-apis`
3. `https://www.postgresql.org/`
4. `https://www.mysql.com/`
5. `https://mariadb.com/`
6. `https://en.wikipedia.org/wiki/SQL_Server_Express`