DSCI 551 - Project Guideline

Spring 2024

Consider a scenario where you need to manage a large amount of data. The size of data is so large that you could not store them in one database. Instead, you partition the data (recall your homework 1) so that some data is stored in one database, while others in another database.

In this project, you are asked to build such a **distributed database** where actual data is stored in SQL or NoSQL databases.

Requirements:

- You should develop a user interface (for database managers) that allows users to insert, delete, and modify the data in the database. For example, your insert function should send the data to a proper database (e.g., according to the hash value of the partition key).
- You should also develop an application [web based](for end users) that uses the database to manage its data. It is up to you to decide what your application is supposed to do and what data it is managing.(Eg. Library Management) [Refer to some resources provided below].
- You should form a group of up to 3 people. Note if you decide to do the project alone, the requirement and expectation will be the same.

Project deliverables and Timelines:

- **Proposal** (due 2/2, Friday, 10 points): project topic, design of your distributed database (how it will partition and store the data), data set to be used for your application. Group members and their roles.
- **Midterm progress report** (due 3/1, Friday, 5 points): tell us your progress so far and the challenges you might have encountered.
- **Demo** (in-class, 4/17, Wednesday, 5 points): Give a live demo of your project. All project members should be present during the demo.
- **Final report** (due 5/3, Friday, 10 points): the final report should be comprehensive, detailing your design and implementation.
- **Implementation** (due 4/17, Wednesday, 70 points): note your project should be fully implemented before the demo. You should include in your final report a link to Google drive where you will upload your project codebase and documentations. Make sure you give access to your project folder.

**** Submission dropboxes will be opened in D2L when the deadlines approach.**

Resources continued on the next page …..

Resources:

Programming Languages: ***Python***, Java, C, C++

Python Databases connectivity:

- https://www.ropstam.com/best-python-libraries-for-database-management/
- https://www.geeksforgeeks.org/how-to-connect-python-with-sql-database/
- https://realpython.com/introduction-to-mongodb-and-python/

Web Applications Development:

- **Frameworks:** Flask, Fast-API, Streamlit, Django etc..
- **[Streamlit] Open Source library for web app :** https://github.com/streamlit/streamlit

** Resources are just suggestions, you are always welcome to use any frameworks, libraries to accomplish the tasks. Choose your teammates wisely, An efficient team will have members from different backgrounds and diverse skill sets. Individual team member contribution is required for this project and we will constantly monitor the progress throughout the semester.