SOEN 363: Data Systems for Software Engineers

Database Project - Phase II, Fall 2024

November 6, 2024

Date posted: Thursday, November 7th, 2024.

Date due: Sunday, December 1st, 2024, by 23:59 *.

Presentation due: Thursday, December 12th, 2024, by NOON *.

Report due: Monday, December 16th, 2024, by 23:59 *.

Weight: 10% of the overall grade.

Group Project. You must work strictly within your group.

Project Outline

This document outlines the phase II of the database project, in which you transfer the data from the relational database you created in Phase I into a NoSQL [1] database. The goal of the project is to make sure the data is completely transferred into the NoSQL so that the performance of the data access as well as query options are compared.

Implementation Platform

You may use any NoSQL database, either cloud-based on available on desktop. Use an programming language of your choice to read the data from the relational database and write the data into the NoSQL database.

Database Size

You database must contain a large data set. While this may be subjective, a database of 300MB would be a good example of a large data set [1].

Design Changes

You may slightly modify the database model while you transfer the data into the new platform. Examples are: removing the IS-A relationships in phase I, or storing Weak entities as part of the data associated within main entities.

Query Implementation

You need to demonstrate the following query types:

- A basic search query on an attribute value.
- A query that provides some aggregate data (i.e. number of entities satisfying a criteria)
- Find top n entities satisfying a criteria, sorted by an attribute.
- Simulate a relational group by query in NoSQL (aggregate per category).
- Build the appropriate indexes for previous queries, report the index creation statement and the query execution time before and after you create the index.
- Demonstrate a full text search. Show the performance improvement by using indexes.

Presentation and Demo

Your project requires a final presentation, during which you demonstrate both databases and compare the data models. The presentation may be 10-15 minutes long. During the presentation you are expected to demonstrate how you populated the data in the relational database as well as how you transferred the data into the NoSQL platform. Include some previews of code snippets. Additionally demonstrate major queries you created in both

[1] see the restrictions that may be imposed by using free NoSQL platforms.

Relational and in the NoSQL database. You may compare the data models and the changes in the two platforms, if necessary.

Demo Schedules

You may decide to demo your project live or submit an off-line video. Live presentations are to be scheduled during the Tutorial time on Monday, December 2nd, 2024. See the submission notes for further information.

Peer Review

Each team is required to peer-review a few other teams. The peer-review is mandatory and the result of the peer review must be submitted at the same time with the final report. Every member of the team must at least peer-review one other team. Note that the peer review does not affect the grade of the teams that are being reviewed but impacts the reviewing team. The links to the videos will be posted on Moodle, once all videos are uploaded. The rational behind the peer review is to make sure everyone becomes familiar with the projects that are implemented by others, and at the same time, every team compare their presentation quality with other peers.

What to submit

Submit your code (data creation / population) as well as the queries along with the data model of your database. In your submission include a report document that provides an overview of your system as well as the approach you used and possible challenges you faced in populating data.

Separate submission links will be provided for presentation videos, final reports, as well as the peer reviews. To submit your video, please make sure your video file is downloadable. Please note that download link is used by the instructor / teaching assistants, only. A separate link will be created and posted on Moodle for peer review. Please do not send links to youtube, vimeo, or other platforms unless the video can be downloaded. If your video file

is below 250MB, you may directly upload it to Moodle.

Submission Notes

- Submit all project code, script, data, etc. under the project submission box by the due date: Sunday, December 1st, 2024, by 23:59. Moodle only allows you to upload files up to 250MBs. While this may be sufficient for zipped folders, you are encouraged to use a public cloud system (i.e. google, dropbox, etc.), and store your project files there. If you wish to do so, please make sure you create a README file and include the link to your folder in there. Make sure the data file can be accessed. Do not include any sensitive data in your public folder. You may submit the credentials to download the file in the very same README file or in a separate file on Moodle.

- The deadline for live presentations are set to the last day of the classes, which will be on Monday, December 2nd, 2024, during the tutorial time. However, if your preference is to present offline, you may submit your presentation video by Thursday, December 12th, 2024, NOON. A separate link will be available on Moodle.

- The deadline to submit the final report is Monday, December 16th, 2024, by 23:59.

References

1. `https://en.wikipedia.org/wiki/NoSQL`