

CMPT506– Advanced Database - Fall 2016

Project

In this project you will design and build a realistic application using a Relational Database and a NoSQL solution using the technology that you have reviewed. You can extend an existing implementation or build an application from scratch. The scope of the selected application must be discussed and agreed upon with the instructor. You need to report your work as a research paper containing: (1) your solution design, (2) a comparison between the relational and the NoSQL solutions, (3) guidelines for migrating the schema and the data from your relational-based solution to the selected NoSQL database, and (4) the key lessons learnt from the project. Additionally you need to present your solution, demo your implementation, and lead a class discussion about your work.

The project requirements 1 to 6 related to the Relational Database solution should be completed and submitted by **4pm on Sunday 13th November 2016**.

The complete project implementation, the research paper and the presentation must be submitted to Blackboard as one zip file by **4pm on Sunday 18th December 2016**. The presentations will be on the same day.

Additionally, mandatory feedback meetings to discuss your interim solution with the instructor are scheduled weekly during the office hours as per agreed upon meetings schedule.

Project Requirements:

1. Document at least 6 **significant** use cases (~2 use cases per team member) that the selected application should support (only consider important use cases and ignore simple ones such as Login). You should include a **use case diagram** and a table briefly describing each use case.
2. Briefly discuss the reasons why the selected NoSQL technology will be suitable for the application. Briefly highlight the key expected benefits compared to a relational solution. The discussion should be specific to your application and not general.
3. Design the Relational Database Model for the application. You should include the ER diagram and the SQL scripts to create database objects.
4. Write scripts to insert test data to test the queries.
5. Write and test the read/write SQL queries to meet the use cases.
6. Define appropriate indexes for your Relational database to improve query performance and discuss the rationale for your choices.
7. Design the NoSQL Database model. You should include relevant diagrams and the scripts to create database objects.
8. Write scripts to migrate the test data used for task 4 to the NoSQL database.
9. Write and test the read/write NoSQL queries to meet the use cases.
10. Define appropriate indexes for the NoSQL database to improve query performance and discuss the rationale for your choices.
11. Conduct and document comparison of the two solutions in terms of complexity of the solution, maintainability, extensibility, etc.
12. Design and run an experimental evaluation to compare the performance of both implementations. You need to design and run experiments to quantify the execution time of the main queries under different workloads and with different database sizes. You might consider using a workload generator such as Apache jMeter to run the experiments and collect performance metrics. You need to present the results using graphs and tables then interpret the results.
13. Document concrete guidelines how to migrate your relational-based solution (schema and data) to NoSQL

14. Discuss lessons learned
15. Present future work highlighting remaining or new questions/problems to be solved
16. Draw key conclusions

The above should be reported in a form of a Research Paper, and then a presentation and a demo should be presented in class.

Ideally your research paper should aim towards the depth and the quality of the paper titled '*vldb2012-Can the Elephants Handle the NoSQL Onslaught*', the pdf of the later posted in Blackboard. Please take a look at this paper and take it as a model to help you shape your work.

This document is a **draft** and further details will be provided via email, discussions and updates to this document.

Grading Schema

Evaluation criteria	%
Depth of the conducted work and adequate level of detail: The more in-depth and specific the better. Clear, concise and accurate reporting of findings. Emphasis database related issues: Abstract and introduction (2 pts) 1. Use cases diagram and a table briefly describing each use case (3 pts) 2. Briefly discuss the reasons why the selected NoSQL technology will be suitable for the application. Briefly highlight the key expected benefits compared to a relational solution. The discussion should be specific to your application and not general (3 pts) 3. Design the Relational Database Model for the application (6 pts) 4. Scripts to insert test data to test the queries (2pts) 5. Read/write SQL queries to meet the use cases (4 pts) 6. Define appropriate indexes for your Relational database to improve query performance and discuss the rational for your choices (4 pts) 7. Design the NoSQL Database model (8 pts) 8. Scripts to migrate the test data used for task 4 to the NoSQL database (3pts) 9. Read/write NoSQL queries to meet the use cases (8 pts) 10. Define appropriate indexes for the NoSQL database to improve query performance and discuss the rational for your choices (5 pts). 11. Comparison of the two solutions in terms of complexity of the solution, maintainability, extensibility, etc (8pts) 12. Experimental evaluation to compare the performance of both implementations (8 pts) 13. Concrete guidelines how to migrate your relational-based solution (schema and data) to NoSQL (8 pts) 14. Discuss lessons learned (6 pts) 15. Future work highlighting remaining or new questions/problems to be solved (4 pts) 16. Conclusions justified and logically sound (3pts)	85
Writing quality and organization of the paper: - Good paper organization with logical order and transitions. - Appropriate writing style, grammar, spelling and proper citation of references . - Appropriate and correct use of technical terms. - All figures are accurate, consistent with the text, and of good quality. They enhance the understanding of the text. Also figures are referenced in the text.	5

Presentation Content & Organization - Presentation provides pertinent, concise and clearly explained information - Material is covered with adequate depth - Presentation well-organized: information presented in logical and interesting sequence that the audience can easily follow - Good quality and neat visual aids	5
Presentation Delivery: - Engaging talk with high level of confidence and enthusiasm - Speaks clearly and uses appropriate language - Meets time limit (30 minutes) - Adequately answering the questions	5
Total	100