Project

This should be completed in groups of 6-8. There will be a presentation and a final report for the project.

Suppose you are engaged as a database designer for an organization to support its operations. You have first to choose and determine the exact nature of the organization. You may pick one that you have encountered in real-life or one that you are interested to work for after graduation (e.g., university, sports association, retail, manufacturing). You have to make certain assumptions concerning the organization's operations, as well as the underlying entities, attributes and relationships. You should state and justify these assumptions clearly, and these assumptions should be reasonable and realistic. Your project must include the following.

1. Analyze the requirements of the organization
2. Identify the relevant entities, attributes, and relationships together with any constraints and properties
3. Produce an E-R diagram for the database
4. Convert the E-R diagrams to relational schemas (clearly indicating the primary keys, foreign keys, functional and/or multivalued dependencies, as well as justifying that your designs are in good, normalized form)
5. Populate the schemas with a reasonable amount of realistic data
6. Produce sample SQL queries on these relations that are used for practical daily operations and activities
7. Produce sample SQL queries on these relations which are of an analytic or data mining nature (this part is optional and carries extra bonus points of up to 5% of the total project mark)
8. Suggest which data fields of the relational schemas should be indexed or hashed, and explain your decision
9. Implement 4 to 6 (and possibly 7) of the above. A good web design will carry extra bonus points of up to 5% of the total project mark.

Additional Bonus (10% points)

Given the swift advancements in Large Language Model (LLM), we encourage you to delve into the application of LLM within the database domain. An additional bonus of 10% will be awarded for incorporating this approach. Your objectives should include:

- Engage with LLM to refine your database architecture. For instance, leverage LLM to enhance aspects such as data integrity (through polishing and validating database schemas, suggesting integrity constraints, detecting anomaly records etc.), independence, and efficient data access in your database design. Document and present a comparison between the newly refined database and your original design. (5% points)
- Investigate methods for crafting prompts that guide LLM to accurately generate queries for extracting information from the database. (5% points)

This project should be completed by 22 April 2024. Project presentations will be held on 23 and 25 April during the lectures. You should prepare a short presentation of 10 minutes (including Q & A). You should strictly adhere to the allotted time and do not overrun. In addition, you should submit a project report by **11 pm 8 May 2024**. The project report should be submitted via Blackboard. Only one submission is required per group. You may submit your report before or after your presentation.

The entire project will account for **15%** of the total mark of the course, with 5% allocated to the presentation, and 10% to the report. With the extra bonus points, the total project mark will not exceed **18%**. The report should indicate the nature of contribution of each member, including the division of labor. However, no member should be unaware of the work of other members, and each member must have a good overall knowledge of the entire project. Good, grammatically correct English is the mark of a good report, and this will be taken into consideration in the assessment of the report.

All members should take part in the presentation. For each group, you should appoint a Group Coordinator, who is responsible for coordinating the various activities and the submission of the final report. Attending the presentation of other groups is not mandatory but is strongly recommended, and you may also interact by asking questions.

By 15 April 2024, each Group Coordinator should tell the lead TA（PAN Zibin) **via a**

provided website link (will be announced soon in the blackboard) to inform him the members in your group together with the title and abstract (i.e., summary) of the project. He will produce a presentation schedule by 20 April 2024.

In your report, you should include the following main sections:

1. Introduction and motivation
2. Design and implementation
3. Conclusion and self-evaluation
4. References
5. Appendices

Additional sections or subsections may be included as appropriate. Please feel free to use your judgment concerning the report organization, and to include diagrams and other illustrations. It is expected that the main report (excluding Appendices) should be no more than around 12 pages. The presentation slides, together with any codes or relevant files, should be included as appendices in the report or submitted as additional files.