**Study Points Assignment on Databases SOFT 2024**

**Objectives**

The objective of this assignment is to enable practical experience in working on a real-life globally important problem, creating programming solution, where design, development, and implementation of a database is in focus. Additionally, the suggested project is intended as a platform for analysis and comparing various database types, their advantages and disadvantages, as well as for creating argumented recommendations about their usability and performance in different domain areas.

You get one project assignment, but are expected to provide three solutions of it:

1. SQL database solution with MSSQL
2. Document database solution with MongoDB
3. Graph database solution with Neo4j

We provide [five datasets](https://cphbusiness.mrooms.net/pluginfile.php/1021087/mod_page/content/7/DBData.zip?time=1707997114296) with original data about the human and environmental protection issues in multiple major cities around the world. The data is in CSV format, downloaded from the CDP Cities, States and Regions Open Data Portal: <https://data.cdp.net/>

It is recommended to extend the data with other public sources, which contribute to creating better picture of the domain area, such as statistics, media publications, glossary of the used terms, etc.

**Tasks**

Your tasks are following:

1. Explore the data and formulate considerations about a hosting database.
2. Design and develop proper database structure of the requested type.
3. Ingest the data into the database, include pre-processing of it, if necessary.
4. Design and develop operations for maintenance of the database.
5. Formulate ten relevant questions for extracting information from the database, design and develop database functionality for implementing the information extraction (for the relevance consult the instructor).
6. Design and implement a model for scaling the database, considering ACID and/or CAP theorem rules.
7. Validate and test all database operations.
8. Evaluate the database’s performance and suggest measures for improving it.
9. Document your work, describing the product and the process. Apply graphical notation (diagrams), when it is possible.
10. Formulate conclusions and recommendations.

**Notes**

This is a group assignment. The optimal size of the groups is 2-3 students.

Solutions bring study points, as follows: SQL DB – 30; Document DB – 20; Graph DB – 30.   
Submission deadlines and locations will be announced additionally by the instructor.