CS306 Database Systems
2022-2023 Spring Semester Group Project
Step 1
Deadline: 15.03.2013 23:59

Welcome to the first step of the group project! Please follow the instructions below to get started.

To begin, please find your project group from the following link.
CS 306 Project Groups (sabanciuniv.edu)

If you cannot find your name in any group, or if you are a new student from the add-drop period, or if any of your group members dropped the course, please contact your TA, Semih Yılmaz, at yilmazsemih@sabanciuniv.edu.

It is important to note that at each step, you need to submit a PDF file from the suCourse. One group member can do the submission. At the beginning of each submission, please clearly state your group name and the names and IDs of all participating group members. If any of your group members do not join the workshops and work harmoniously, you may not submit their names for that project with a consensus decision. The group project grading will be calculated based on your submission at each step.

This project can be a valuable addition to your portfolio and can be used for job applications. You can use your Github repositories to showcase your work to potential employers. I am certain that this project will help you in your future endeavors. Therefore, I would like you to create a private repository on GitHub and share it with me and our TA, Semih and Hasan. Please keep all your project files in the repository and share the link with us in your PDF submissions along with your explanations. We will review the work you commit before the due date. Please note that you are responsible for uploading your work before the deadline and ensuring that it is not deleted by mistake. There will be no late submission. Please be careful at your work and deadline!

1. For your first step, visit the website https://ourworldindata.org/ and choose a topic of interest that is not related to Covid-19, as we will be studying the Covid-19 dataset during recitations.

2. The complexity of your project will be proportional to the group size. You will need to establish tables that include at least the same number of group members besides location tables in the later steps of the project. If you are unable to form enough tables with your chosen topic, you may use other datasets and join them with the location data. For this reason, I suggest that you choose a dataset that includes location data (countries).

3. You need to separate the dataset from the flat design into smaller portions based on the entity sets, the relationship sets and the attributes. You do not need to keep all the columns; please specify the most important information for you and just keep those attributes. Please remember what we did at the recitations for the Covid-19 dataset. You can follow the same steps for your datasets. Please clean and remove duplicates from your final sets, create your CSV files and upload to your repository.

   Based on our work, you need to differentiate your entity sets, relationship sets, and attributes. Create an ER diagram and upload to your repository. Please identify and explain the constraints and custom constructs if you have any that we learned from the lectures and recitations. You may want to create a readme file for the explanations in your repository or upload the same PDF file that you are going to submit at suCourse.

4. In your PDF file, provide a project description that includes the title of your project. In your description, specify what your database application will be, what data you will store, and which world problems you are addressing. You need to include your ER diagram and all the explanations from the previous step. Identify your CSV files and how you separated them from the flat design. Please briefly explain the cleaning steps you took, such as removing duplicates from your final sets, etc. Provide a link to your repository.

If you have any questions or need further clarification, don't hesitate to reach out to me or your TA Semih.

I wish you all the best of luck with your project!