Assignments

The problem/the business case for Assignments 1 to 3

Scenario 1: universal math-education (or a free math-enrichment program on the web)

A non-profit organization interested in universal math-education (or a math-enrichment program) wants to develop a website which allows any child in the world who has access to the internet to do math home-work practice using their math worksheets. Initially, they want to start with math material related to K1-12 and later expand to cover practice material for college level as well. The students will be able to learn math in this manner because at the beginning of each math worksheet there would be a brief introduction to the math concepts covered in the particular sheet or links to sections of text books freely available on the internet where the relevant concepts are explained. What is only required from the learner is the motivation for self-learning.

This organization has a pool of voluntary teachers who are willing to mark the work sheets done by the students and provide feedback. Initially it will be only available in English, but later they want to translate the math work sheets to other world languages as they find volunteers to do the work. What a student will do?

A student who access this website for the first time needs to register herself. Then she can start at a level which is appropriate for her (as she thinks). Download the worksheet, attempt questions, make a scan or take a clear picture of the finished work with a mobile phone and upload her work for marking. Within a specified time, a volunteer teacher will mark her work and upload the marked assignment with the teacher's comments. Then the student has either the option of redoing it (or attempting a similar math worksheet) or go to the previous/next math work-sheet as suggested by the teacher. So, the process will continue that way as the student progresses with her learning.

The organization wants to store contact details, the level at which the student is studying in her home country and keep track of the progress of all students registered with the website (i.e. the level of the math worksheet, marks allocated, teacher's comments, etc.) because they want the best students to be trained for Math Olympiads (using the same process as for home-work training), keep details of volunteer teachers (their contact details, availability, languages in which they could provide feedback and expertise), details of all math worksheets (level, created by whom categorized by the field; geometry, algebra, etc.) and to store math worksheets.

Further, depending on the available resources they hope to have Skype (or similar video conferencing) sessions between a student and a teacher in the student's native language or in English (as required).

Scenario 2: Academic learning using freely available material on the internet.

A non-profit organization wants to develop a website which will assist their users in online learning by using freely available material on the internet. The users who want to use the website need to register themselves before accessing the material on site/make their recommendations.

Develop a database for this organization's website which assists in online learning by giving the sequence of learning material (academic or otherwise: freely downloadable books/academic articles,

etc.) for a beginner to the professional in a specific subject area. Should give links to online material (available options: MOOCs, course material of top universities, textbooks used in each class, topics covered and to what depth, available homework/practice or exam material, etc.) and store the sequence in which that material should be followed. The original material should be stored by produced date or when the course/MOOC was offered, and users should be able to query the database by institution, topic, date, MOOC/other academic course, view the material of their choice and store their preferred possible sequences (there might be a few) in which the material could be followed. The sequence to follow for learning/to gain competence could be from different courses (but valid sequences as defined by the user according to taste, difficulty, etc.). other users visiting the website should be able to view these stored sequences recommended by other learners with their comments and store their own recommended sequences. Users should be able to recommend links to original course material available on the web; which they think are suitable for a specific course, if that link is already not listed on the website.

You could start thinking of what material to include in an introductory database academic course (like CSC 370) and the sequence(s) in which the material should be followed. Expand it to other courses say, in computer science and later to other courses in engineering/science.

Instructions for assignments 1 to 3:

For **one** of above given scenarios, develop the necessary documents and finally implement you design. The extent of work (tentative) that should be done for each assignment is specified below:

As you read through the given scenarios, you will understand that this is not a detailed requirement specification. You might have to find out their requirements to the extent possible and finalize on the precise requirements.

You are supposed to develop a database to handle the current and future requirements of the respective organization's website.

1st assignment

Develop the business case further and **write your assumptions**. Strat with the ER modelling process

Determine the main objectives of the system, identify tasks performed by different users during a typical day, what type of data will be associated with these tasks, define the scope of project and the relevant data, think of possible outputs with the given data and come up with the initial data model (from the data identified in this step; what could be entities, attributes and relations?)

Use the 'Entity Relationship Diagrams with draw.io' at https://about.draw.io/entity-relationship-diagrams-with-draw-io/ or similar package to draw the ER diagrams. Or use UML (Universal Modeling Language) notation as shown at UML class diagram. Also, you can check the following link from Quora for more free options on drawing diagrams:

What is the best free DB schema design tool? - Quora

2nd assignment

Review the data model developed in Assignment 1. Try to introduce a new entity where you have many to many relationships. Think about the values of optionality and cardinality. Check for inheritance. Finalize ER diagrams and start on the relational model (refine your data model if required). Identify potential problems in the database schema and refine it. Define the domains for all attributes. Add some constraints on the data values if necessary. Identify primary and foreign keys. Create tables, load some data and run some queries on the developed relational model. State your assumptions.

3rd assignment

Identify functional dependencies, check for poorly structured tables having problems with updating data (modification, insertion and deletion) problems. Check whether your tables are properly normalized? Load some data and run some queries on the developed relational model (refine your data model if required). Define some user views. State your assumptions.

Implement the database using PostgreSQL, load some data and try to connect via an application programming language interface to do some updates/queries on the data in the database

4th assignment (on Google BigQuery)

After familiarizing yourself with Google BigQuery select a dataset and examine the schemas, descriptions and data tables of your selected dataset. This is to get an understanding of the data that you will be working with for the assignment. You can run some simple queries over the tables to get a feel of them or use BigQuery's 'Preview' tab to see what the data looks like. Next, try asking some meaningful questions about the dataset (you can use your experience with example queries on the sample datasets). Next, try to translate a question in plain English to a schema or we want to read the tables and explore the data and think about which tables and columns are necessary in answering a question we are asking. This skill is both necessary and is exactly how real-world data querying and analysis works!

Your queries should be fairly efficient -- they should each take at most ten seconds to execute on BigQuery, and most of them will be finished in less than ~4 seconds. If any of your queries are taking much longer than that, you've probably written them in particularly inefficient way; please try rewriting them. Please also check to make sure you're not querying more than a couple GBs of data to fit within the 1TB of free querying.

You can save your queries for each question from the BigQuery interface directly, or you can keep track of your queries in separate files yourself. Remember that you can use BigQuery's "Query History" tab to inspect previous queries you've run. Write standard SQL queries to answer you questions. Try to think about ten questions and write appropriate queries.

Take a snapshot of your query and the result as seen on BigQuery for all ten questions. Paste them into your journal with the related ten questions (in English) and with any other explanatory notes related to each question/query

All assignments will be done in groups (but try to contribute your own part while learning from others) and maintain a journal about your contributions, which must be uploaded at the set deadline for each assignment.