

Lab 2: Database requirements analysis and design
Section 01

Due date: January 26, 11:59pm.

Note: This lab involves a number of deliverables that are due before January 26. The due date above is for the final set of the deliverables.

This is a **team lab**. Each team submits one set of deliverables. Each team member receives the same grade for the assignment.

Course Project Information

This quarter, CPE 366 comes with a group course project. General information about the project is provided below.

Customer. Our customer for this quarter is Dmytro Marushkevich from the San Luis Obispo office of Sapient Razorfish (formerly known as Rosetta), a marketing agency with an office in San Luis Obispo.

Project. The marketing campaigns run by Sapient Razorfish (Rosetta) and decision-making on behalf of its clients are always data-driven. As part of the project, you are being asked to put together a data warehouse containing the information about a customer of Sapient Razorfish, and develop software that supports some of the decision-making activities of the company's business analysts.

Lab Overview

In the course of this lab, you will

- receive an overview of the customer requirements via direct communication with the customer (January 17);

- elicit additional requirements for the database via direct communication with the customer (January 19);
- work on the database design, and create an Entity-Relationship model for the project database (January 19-26).

The lab assignment spans five lab periods, and has three distinct steps, each of which serves a different goal and yields a set of deliverables. This document specifies your assignment for each of the periods as well as final deliverable.

The overview of the lab structure is below.

January 17, 2018: 9:10am - 10:00am: Groups are announced. Dmytro Marushkevich joins us for an initial presentation. He will discuss the company, its business model and will give a brief and broad overview of the desired database and software application. We release the project documentation to you.

January 19, 2018: 9:10 - 10:00am: Q&A session with the customer. Each team will have an allocated portion of time for questions. Questions/answers must be recorded and posted later.

January 19 - January 26, 2018: teams work on the database design.

Assignment

January 17: Group Creation and Initial Presentation

I will announce the group lineup at the beginning of the class on January 17. Teams will contain three to four people.

Following team formation, our customer will present an overview of your project, followed by a short, informal Q&A session. The main purpose of the presentation is to give you a good understanding of what the project entails. After the presentation, we will release customer-supplied documentation which will contain two key pieces of information for you:

- Data description.
- Software use cases/requirements.

This document, combined with the data, which will be released upon your completion of Lab 2 will serve as the basis for your project.

Assignment. Attendance is required for everyone. During the presentation, you are welcome to take notes - those will come in handy later. Upon receiving customer documentation, study it in preparation for the April 19 Q&A session.

Deliverables. Each team shall come up with a name. Each team will send one email to me with

- the name of the team;
- the list of github userIDs for all team members;
- the list of gmail accounts *or email addresses that you want to use for sharing Googledocs files* for all team members.

January 19: Question and Answer Session

Preparation. Before the *January 19* meeting with the customer, each team shall study the materials provided by the customer. The goal of each team is to identify any aspects of the application domain, information about objects to be stored in the database, customer needs (w.r.t. the database), *that require additional clarification.*

Each team shall prepare a list of questions that the team wants to ask the customer during the Q&A session. Note, that in addition to the customer, you may direct your questions at the instructor.

Please note: initial list of questions **must be prepared** by each team independently of other teams. However, once lists are prepared, I allow for teams to compare them. I expect that many questions will coincide or be similar (it is ok if multiple groups have similar/same questions). It is also ok for one group to come up with a question, no other group has asked — there is no need for other groups to add it to their lists, since each question only needs to be asked once, and all teams benefit from the answer.

Each team shall create a google doc of its questions to the customer named CSC366-<TeamName>-Questions-to-customer and share it with dekhtyar@gmail.com (please note – share all documents with my Gmail address!!) **before** the class time on Friday, January 19.

Q&A session. Each team will be given 10-12 minutes of time to ask their questions and receive answers from the customer and the instructor. We will set the order of questioning by a simple lottery at the beginning of the lab. I strongly suggest that each team sits and works together during the Q&A. Please note the following:

- Listen carefully to the questions other groups are asking. It is very likely, that some of your team's questions will be posed by teams who get to ask questions earlier.
- If your team runs out of questions yield your time to the next group.
- If your team has unanswered questions left after its Q&A period, wait until all groups ask their questions. If the question has still not been asked, you will have a chance to ask it at the end of the lab.
- You are certainly allowed to ask questions that are not on your list - some questions might occur to you as a result of things said during the Q&A session.

Each team is responsible for recording the answers to all questions its members have asked (whether those questions were prepared or were asked on the spot).

After each team had a chance to ask questions, if any additional questions remain, a free-form Q&A period will begin and will last until the end of

the lab period. Any questions that went unasked, or unanswered during the Q&A session should be submitted in writing (via email) to the instructor. The questions will be forwarded to the customer, who, in turn, will provide answers prior to the Monday, April 13 lab.

Deliverables. After the Q&A session, all teams, in collaboration shall create the project knowledge base that consists of the *thematically organized* Question-Answer pairs. The exact structure of the knowledge base is left up to the teams (generally speaking, you want to break the entire knowledge base by topic: e.g., group information about one type of data in one place, and information about another type of data — in another). A single question can appear in multiple places in the knowledge base (as it may be applicable to more than one "leaf" topic).

The knowledge base shall be stored as a Googledocs file created by the instructor ahead of January 19, and shared with all of you. You are responsible collectively for making this document both correct (i.e., reflective of the actual answers you receive from the customer) and complete (i.e., making certain you did not miss documenting any Question-Answer pairs).

The knowledge base is one of the two **main deliverables** for this lab. The initial state of the knowledge base shall be completed by the **main Lab 2 due date**, however, you should continue using and maintaining it throughout the entire quarter.

January 19 - January 26: Database Design

Assignment. The main task of this lab is for each team to prepare an initial full database design document.

During January 22 and January 24 lab sessions, each will will work with the instructor on building the E-R model for the project database. Depending on the level of progress we may conduct one or two structured activities at the beginning of each lab period. Alternatively the entire lab period may be given to the teams for team meeting activities, with instructor participating in each team's meeting in turn.

After the *January 26* lab, each team shall have a finalized initial conceptual model for the project database.

Deliverable. Each team is given time until the end of the day **Friday, January 26, midnight** to finalize and submit the database model document.

The design document prepared and submitted by each team shall, at a minimum, contain the following information:

- List of entity sets for the proposed database.
- List of attributes for each proposed entity set.
- Identification of primary keys for each proposed entity set.
- Identification of all weak entity sets in the proposed database, and of their discriminating attributes.
- Identification of any other entity set constraints for the proposed entity sets.
- List of relationship sets for the proposed database. For each relationship set, the following must be indicated:
 - All participating entity sets.
 - All clarifying relationship set attributes.
 - Type of the relationship set (one-to-one, one-to-many, many-to-many).
 - If the relationship set is an identifying one for a weak entity set.
- List of other relationship set constraints.
- List of any class hierarchies and/or aggregates in the database model.
- List of relationship sets (with all the information as above) associated with the aggregates.
- E-R diagram of the proposed database design.

The design document must be typeset (handwritten submissions will not be accepted). See comments above about the software to use to draw the E-R diagram. The E-R diagram shall contain all entity sets and relationship sets. It shall also contain all key attributes for all entity sets and all relationship set attributes. Other attributes can be omitted from the diagram for clarity. All constraints that can be shown on the diagram, shall be shown.

The design document shall begin with the name of the group, and the list of group members. Note, that in general, **all documents** submitted by each team during the course of the project must contain the team name and the list of students.

Deliverables Overview and Submission Instructions

Overall, the lab has the following formal deliverables:

No.	Deliverable	Due:
1	List of questions for Q&A session	January 19, 9:10am
2	Project knowledge base	January 26, midnight
4.	Database Design document	January 26, midnight

Submission. All deliverables shall be submitted as googledocs. For the database design, each team shall create a googledocs file <TeamName>-Lab2-Design and share this file with dekhtyar@gmail.com. The sharing of the file shall happen after the document is completed, and will be considered the act of submission.

Please, keep the soft copies of all submitted documents. You will be working with them in the labs that follow. In fact, make sure that you retain a softcopy of each submission. Lab 3 will require you to make changes to your Lab 2 submissions. You should make a copy of all Lab 2 artifacts (except for the project knowledge base), prior to editing them for Lab 3¹.

Grading. While you may not receive the numeric score for your submission until after Lab 3 is graded, you will receive detailed feedback on your database design by Monday, January 29.

¹Strictly speaking, we could engage in some straightforward version control, but for the class, I am experimenting with the use of Googledocs as the mechanism for assignment submissions, because I believe that I can actually be more proactive in grading work submitted this way.