

Concordia University
Department of Computer Science and Software
Engineering

SOEN 363: Data Systems for Software Engineers – Fall 2025

Project

(Phase 1)

Posted Date: September 20, 2025.

Due Date: October 29, 2025, at 11:59 PM.

Evaluation: 10% of final mark.

Late Submission: Please check the course outline.

Type: Group project.

Purpose: The purpose of this phase of the project is to design and implement a Database.

CEAB Attributes: This phase of the project is primarily evaluating your use of engineering tools.

1 Project description

1.1 Overview

Every person who enters a hospital to receive healthcare services is known as a patient. The patient is first registered with a unique identity that remains with them for life. This identity connects all of their medical information, whether it's a hospital visit, an ICU stay, a doctor's note, or a diagnosis. Basic details such as date of birth, gender, and life status are recorded once and remain the same across their medical history. For privacy, certain details, like the birthdays of elderly patients, may be adjusted while still preserving accuracy.

When someone arrives for care, their hospital visit is documented as a distinct event. Each visit has its own identifier, marking when they were admitted, when they left, and—if it occurred—whether they passed away during that stay. The record also notes the nature of their visit, such as emergency, elective surgery, urgent treatment, or even newborn care. Additional information may include how they arrived (for example, through the emergency department), their insurance, marital status, or other demographic details. Together, these admission records provide a structured history of every patient's encounters with the hospital.

Each hospital has several wards, e.g., the intensive care unit (ICU). For those who require critical care, their time in the ICU is closely monitored and tracked in detail. Each ICU stay has its own unique identifier, which may encompass multiple units if the patient is

transferred between them within a short timeframe. These records capture the date and time the patient entered the ICU, the date and time they left, the type of ICU they were in (such as medical, surgical, or cardiac), and even the physical ward location.

Throughout their stay, clinicians produce a wide range of notes documenting the patient's journey. These include discharge summaries, radiology reports, ECG and echocardiography interpretations, progress notes, and more. Each note is connected to the patient who was admitted to the hospital on a specific date and time. Notes also record who authored them and whether they contain any flagged errors. Taken together, they form a rich, narrative record of the patient's care.

At the conclusion of a hospital stay, formal diagnoses are assigned using the International Classification of Diseases (ICD) code. These codes are assigned to each patient that admitted to a hospital at a specific date and time by professionals in the healthcare system and capture the conditions treated. Each diagnosis links back to the patient and their hospital stay, while the codes themselves are defined in an official medical dictionary that ensures consistency across the hospital system. For example, "401.9" in ICD code version 9 is "Unspecified essential hypertension".

In summary, a patient may have multiple hospital visits across their lifetime. Each visit may include zero or more ICU stays, multiple sets of clinical notes, and a list of diagnoses defined by standard codes. Together, these elements trace the patient's journey through the healthcare system—from registration, through treatment and documentation, to final outcomes.

Summary of Relationships:

- PATIENTS → ADMISSIONS: A patient can have multiple hospital admissions.
- ADMISSIONS → ICUSTAYS: A hospital admission may include zero or more ICU stays.
- ADMISSIONS → NOTEEVENTS: Notes are associated with a particular admission.
- ICUSTAYS → NOTEEVENTS: Notes can also be linked to ICU stays.

ADMISSIONS → DIAGNOSES_ICD → D_ICD_DIAGNOSES: Each admission can have multiple diagnoses, and each diagnosis is defined in the ICD dictionary table.

This narrative forms the backbone of the hospital database, capturing a patient's full journey—from admission, through ICU care if needed, to clinical documentation and final diagnoses.

1.2 Data sources

Please use the MIMIC-III dataset as your primary data source, available at the following link: [MIMIC-III Documentation](#). Read the documents related to the database. For the purpose of this project, only the following tables need to be read: **PATIENTS**, **ADMISSIONS**, **NOTEEVENTS**, **DIAGNOSES_ICD**, **D_ICD_DIAGNOSES**, and **ICUSTAYS**. The information related to each table is also available here: [MIMIC-III Tables](#). The remaining

tables are not essential and should be excluded to keep your database focused and manageable. The link for download the data for the MIMIC-III dataset is available here: [MIMIC-III Data](#).

1.3 Project phases

This project consists of two phases:

- Phase I addresses the relational database. In this phase, you will design, implement, and populate the relational database from the mentioned scenario.
- Note that the output of phase I will be used to populate the NoSQL database in Phase II. While you submit phase I of your project now, during phase II, you may enhance and refine your database to address unforeseen design issues. The final presentation will cover both phases.

1.4 Task 1: Drawing ER Diagram

For this task, you are required to carefully read the project overview description, and construct an Entity-Relationship (ER) diagram. Your diagram must include all relevant entities, their attributes, and the relationships among them. Be sure to clearly specify the primary keys and, where applicable, foreign keys within your diagram. The ER diagram must be created using a digital tool of your choice (e.g., Lucidchart, Draw.io, Microsoft Visio, or any other ER modeling software). Handmade or scanned diagrams will not be accepted.

1.5 Task 2: Making the Database

Based on the ER diagram you created in the previous section, you are required to design and create the database. Ensure that all entities, attributes, and relationships from your ER diagram are properly mapped into tables with appropriate primary keys, foreign keys, and constraints. After creating the database schema, you must import the data from the provided data sources into your database. Make sure the data is correctly loaded and consistent with the design of your schema. The link for download the data for the MIMIC-III dataset is available here: [MIMIC-III Data](#).

Note: This

1.6 Task 3: Writing SQL Queries

Using the hospital database schema described earlier (PATIENTS, ADMISSIONS, ICUSTAYS, NOTEEVENTS, DIAGNOSES_ICD, and D_ICD_DIAGNOSES), answer the following questions by writing the appropriate SQL query. For each query, execute it on your database and provide both the SQL command and the resulting output.

1. List all patients (patient IDs) along with their date of birth and gender.

2. Find all admissions for a given patient (use a sample patient ID) showing all the information.
3. Retrieve the number of admissions for each patient.
4. Show all the patients who were discharged to "home".
5. Show the list of all the patients who have private insurance.
6. Show the list of all the patients who transferred from one location to another location in the hospital.
7. Find patients who have more than one ICU stay in a single hospital admission.
8. List all the patients that were in the ICU and the first care unit, and the last care unit are "MICU".
9. Retrieve all notes written for a specific admission ID, including the author and whether an error was flagged.
10. Show the first 10 discharge summaries recorded in the database.
11. Count the number of notes written per admission.
12. List all diagnoses (ICD codes) assigned to a given patient, including the textual description from the ICD dictionary.
13. Find the top 5 most common diagnoses in the hospital database.
14. Retrieve all admissions that include an ICU stay and at least one diagnosis of "hypertension" (ICD-9 code: 401.9).
15. Show all patients who have never been admitted to the ICU.
16. List of patients who had at least one radiology report during admission.
17. List of patients who had at least one radiology report from the chest on admission.
18. List of patients with discharge summary report during the hospitalization.
19. List of patients with radiology report or ECG report during the hospitalization.
20. Generate a summary report showing for a patient, the total number of admissions, the number of ICU stays, and the number of diagnoses recorded.

Note: All the queries must be covered, and based on the size of the team, each student must be in charge of implementing 4 to 5 queries.

2 Submission Guidelines

1. Submit a **PDF document** that includes your ER diagram for Tasks 1 and the Queries and a screenshot of the result for Task 3. In this document, please ensure that you paste your code for each task into the specified query (from Q1 to Q20). Ensure that your file is named according to the following format: `SOEN363_Project_Phase1_YourTeamName.zip`.
2. Submit your assignment via Moodle before the due date.
3. The project must be demonstrated online to the marker at the specified date and time. This schedule will take place within one week after the due date, and further details regarding the specific date and time will be announced later. Make sure to keep an eye out for those updates!

3 Rubric

Table 1: Rubric for Assignment 2.

Task	Description	Points
Task 1	Drawing ER diagram	20 points
	<ul style="list-style-type: none"> All entities, attributes, and relationships included, PKs and FKs clearly defined, and correct cardinality (10pt). Clean, well-labelled diagram made with a digital tool (10pt). 	
Task 2	Creating database and importing the data	20 points
	<ul style="list-style-type: none"> Schema fully matches ER diagram. Tables, PKs, FKs, constraints correct. Data imported accurately and completely. No integrity issues.(20pt). 	
Task 3	SQL Queries	40 points
	<ul style="list-style-type: none"> Each query 2 pt . 	
Task 4	Documentation & Presentation	20 points
	<ul style="list-style-type: none"> Documentation including a clear drawing of the ER diagram. SQL queries include code & output. Professional formatting. (10pt). Presentation (10pt). 	
Total		100 points