Assignment 1 — Databases 2020

Michael Emmerich, Chivany van der Werff, Laduona Dai, Philippe Bors, Luuk Nolden

March 2020

PLEASE READ THE GENERAL INSTRUCTIONS CAREFULLY BEFORE YOU START WITH YOUR SOLUTION. DUE TO THE LARGE SIZE OF THE CLASS, THE ADHERENCE TO THE INSTRUCTIONS IS INCREDIBLY IMPORTANT TO MAKE IT POSSIBLE FOR US TO CORRECT THE ASSIGNMENT SWIFTLY.

Grading will be handled semi-automatically where possible. As a result, your submission should conform to the following rules:

A **TURNING IN YOUR WORK:**
Please deliver your work for the assignment to `db2020-team@lists.liacs.nl` (the submission email will be active from March 15th, 2020). Submit your assignment before or on **April 10th, 19:00**. Start early with the completion of the Assignment, because late submissions will not be accepted.
Please write your Student-IDs and names on each submitted page. Team size can either be 1 or 2. It is not allowed to copy solutions from other teams.

B **SUBMISSION STRUCTURE:**
Send a single .zip called ***<studentno1>_<studentno2>_A1_DBMS.zip*** (depending on the amount of members in your group), where DBMS should be "**oracle**" if you used the oracle/sqlplus database manager, or "**sqlite**" if you used sqlite. For example, students s1403492 and s1639485, who hypothetically speaking used oracle, should hand in a .zip file called ***1403492_1639485_A1_oracle.zip***
The .zip's internal file structure should match that of figure 1. The files in the folders should match the names as in figure 1.

```
studentno1_studentno2_A1_DBMS.zip
├── 01_ER_Modeling
│   └── ERDiagram.pdf
├── 02_SQL_Design
│   └── SQLDDL.sql
└── 03_SQL_Queries
    ├── 01.sql
    ├── 02.sql
    ├── 03.sql
    ├── 04.sql
    ├── 05.sql
    ├── 06.sql
    ├── 07.sql
    ├── 08.sql
    ├── 09.sql
    └── 10.sql
```

Figure 1: **Your submission must match the file structure above.**

- The **ERDiagram.pdf** file should contain one ER diagram. This has to be a digital rendition. You can use tools like draw.io to fulfill this requirement. Scans and/or pictures of hand-written diagrams will not be graded.
- The .sql files for the queries part should contain the answer to the question number in the filename.
  Make sure that your .sql files work ***without errors*** in the DBMS you chose.

C **ASSUMPTIONS:**
Assumptions made for your ER diagram should be included in the aforementioned **ERDiagram.pdf** file. Assumptions made for the SQL queries or SQL design should be included as comments in all relevant .sql files.
Note that assumptions only are approved if they are justifiable.

Failure to comply to any of these points will result in a penalty to your grade.

1. **PLAGIARISM:**
   Because Assignment 1 is part of your Final Grade, it is **strictly prohibited to copy any part of the work of other students**. Violation of this rule will be considered as **plagiarism** and will be sanctioned by **the Board of Examiners (de Examencommissie)**.

2. **RECOMMENDATIONS:**
   Please start early with the Assignment and use the **practicals and lecture breaks to ask questions**, if possible, in order to avoid excessive emails.

3. **UPDATES:**
   All updates concerning the Assignment will be published on Blackboard in the Announcement Section.

1 ER Diagram Modeling - Viruses (40%)

It is mandatory to use the ER diagram style that was discussed in the lectures.
Because of recent Virus outbreaks, the government would like to have a database created to keep track of existing Viruses and corresponding Viral diseases as well as the health of its citizens and the treatment(s) they are getting at the hospital they are recovering in.
For this part of the assignment your goal is to design an ER diagram for this new governmental database that keeps track of the aforementioned elements; The following information for your ER Diagram is provided by the government, virologists, and hospitals:

1. A Virus has a name, an ID, and a date of discovery.
2. A Viral Disease has a name, an ID, and a date of discovery as well.
3. A Viral Disease is caused by a Virus and you can assume that a Virus causes one Viral Disease.
4. Viral Diseases are distinguished between being fatal and nonfatal.
5. Viruses infect. Each citizen in this database has been infected. They, therefore, have at least one Viral Disease.
6. A citizen has a name, a BSN, and an age.
7. Unfortunately, fatal Viruses have killed and are killing people. When that occurs, a time-of-death is noted.
8. Luckily there are treatments available. A treatment has a name, step 1, step 2, step 3, an ID, and an estimated recovery period.
9. Doctors are the only ones that are allowed to perform these procedures. A Doctor has a name, an age, an ID, and a specialization.
10. A Doctor can either be an Attending Physician or a Resident (Google if you are not familiar with those concepts). Residents assist the Attending Physicians in their work.
11. A Doctor works for a hospital.
12. A hospital can have multiple Doctors employed.
13. A hospital has a name, an ID, and a location.
14. A citizen is treated with a specific treatment by either one Doctor or a Doctor and his/her Residents.
15. A Doctor treating a specific citizen can appoint one or more treatments to that citizen.
16. One specific treatment carried out by a Doctor can be appointed to multiple citizens.
17. Each individual hospital has been directed to check up (only within their own hospital) on the processes of citizens getting treated by their doctors in case something goes wrong.
18. Such a process is checked by just the hospital it takes place in.

2 Logical Database Design (E/R to DDL) - Hospital (30%)

In Figure 2 (below) an E/R diagram is given for a hospital. Translate this diagram to SQL CREATE TABLE statements that create the tables for the entities and relationships, enforcing the given constraints as well as possible, if possible by assigning appropriate keys and combining tables if needed. In general, use your best judgement to figure out which data types correspond to which attributes and double-check to see if your work runs without error in your database management system of choice (ORACLE or SQLLite). The IS-A subclasses have no full coverage and overlap is not allowed.

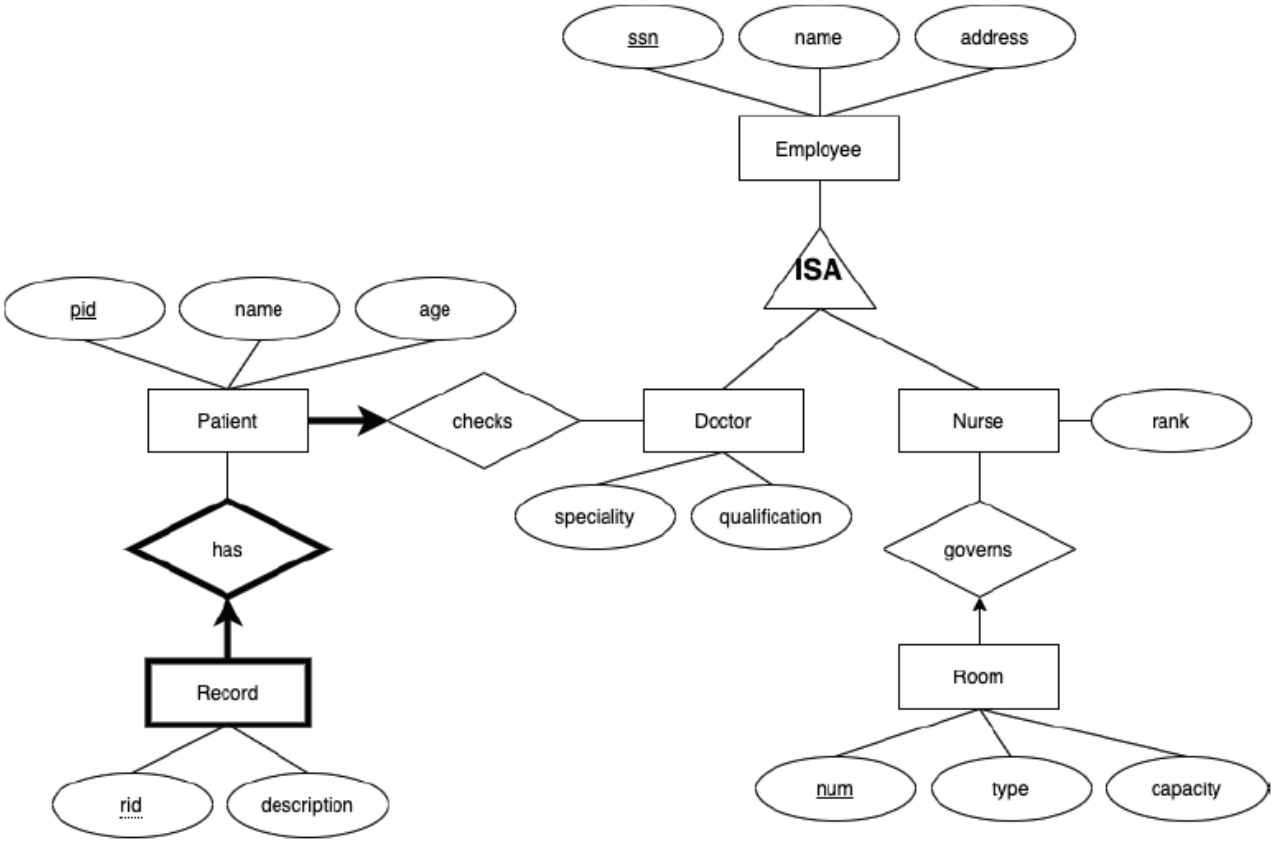

Figure 2: Part 2: E/R diagram to be translated to SQL.

3 SQL Queries - Movie Database (30%)

PART 1.3 – SQL Queries (30%) Download the files MovieDatabase.sqlite.sql and part13.sql from blackboard and program the SQL Queries formulated below:

- Query1: Find the name and age of all actors under the age of 18.
- Query2: Find the name and age of all actors under the age of 18 and over the age of 65 in ascending order.
- Query3: Find the names and a_id of all actors that are action-movie actors.
- Query4: Find the names of all actors that have won an oscar for their own genre.
- Query5: Your little brother wants to see John Fist, but is only 12 years old. Find a movie to watch.[1]
- Query6: Find the names of all actors that have not won an oscar.
- Query7: Return a list of actors and the total profit of their movies.
- Query8: Find all the movie studios that made two or more movies and list the amount. Order Descending
- Query9: For every award year at the oscars, list the most successful genre.
- Query10: Many movies appear per year: find the year in which the most oscars were won.

END OF Assignment 1. Wishing you success!

[1] (for up to 12 years old).